

Univerzita Karlova v Praze
Filozofická fakulta
Ústav informačních studií a knihovnictví

Studijní program: informační studia a knihovnictví

Studijní obor: informační studia a knihovnictví

Mgr. Veronika Synková

Aktuální problémy identifikace a popisu
webovských informačních zdrojů

[Actual problems of identification and description
of web information resources]

Rigorózní práce

Konzultant rigorózní práce PhDr. Eva Bratková

Praha 2007-09-07

Prohlášení:

Prohlašuji, že jsem rigorózní práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje.

V Mratíně, 7. září 2007

.....

podpis rigorózanta

Identifikační záznam

SYNKOVÁ, Veronika. *Aktuální problémy identifikace a popisu webovských informačních zdrojů [Actual problems of identification and description of web information resources]*. Praha, 2007. vi, 161 s., 26 s. příloh. Rigorózní práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví 2006. Konzultant rigorózní práce PhDr. Eva Bratková.

Abstrakt

Práce se věnuje problematice identifikace a popisu elektronických informačních zdrojů na Internetu. Zaměřuje se především na zdroje dostupné prostřednictvím služby WWW. Nastiňuje specifika elektronických zdrojů a důvody, proč se pro jejich popis nepoužívají klasické katalogizační metody, případně jsou pro jejich popis kromě klasických katalogizačních metod použity i metody jiné. Základním cílem práce je přiblížit čtenáři vybrané popisné metadatové formáty a ukázat, jak jsou použity praxí. -- Stručně je popsána problematika volby popisné jednotky a úrovně popisu elektronických zdrojů z pohledu studie IFLA Funkční požadavky na bibliografický záznam (Functional Requirements for bibliographic Record, FRBR, <http://www.ifla.org/VII/s13/wgfrbr/index.htm>). Jsou uvedeny bibliografické popisné jednotky, které rozlišuje tato studie, jejich vzájemné vztahy a vztahy k ostatním entitám, které se týkají bibliografického popisu zdrojů. -- Práce stručně zmiňuje také související technické standardy pro komunikaci metadat - značkovací jazyky, problematiku Rámce pro popis zdrojů (Resource Description Framework, RDF, <http://www.w3.org/RDF/>). Uvádí nově také schéma pro strukturální metadata METS (Metadata Encoding and Transmission Standard, <http://www.loc.gov/standards/mets/>). -- Navazující kapitoly jsou věnovány vybraným popisným metadatovým formátům a systémům a projektům, které na ně navazují. Každý formát a následně systém či projekt jsou stručně popsány a analyzovány z hlediska popisných jednotek, struktury formátu, příp. změn vůči původnímu metadatovému formátu. Vybrány byly metadatové formáty, které se v současnosti používají pro popis elektronických zdrojů nejvíce. Z řady projektů a systémů, které vybrané formáty používají, byly zvoleny takové, na kterých je možné demonstrovat různé způsoby použití původního formátu. -- První z těchto kapitol je zaměřena na metadatový formát Dublin Core (<http://www.dublincore.org/>). Z projektů a systémů, které tento formát používají, byl vybrán australský metadatový formát pro popis vládních informacních zdrojů AGLS (Australian Government Locator Service, <http://www.agls.gov.au/>) a australská síť pro vzdělávání EdNA (Education Network Australia, <http://www.edna.edu.au/>). Na formátu ETD-MS, který je používán pro Digitální knihovnu diplomových a disertačních prací (NDLTD, <http://www.ndltd.org/>), je ukázáno použití metadatového formátu Dublin Core pro popis vysokoškolských prací. Do této kapitoly patří také české projekty WebArchiv (<http://www.webarchiv.cz/>) a ANL FULL (<http://full.nkp.cz/>). -- Následující kapitola pojednává o metadatovém formátu TEI (Text Encoding Initiative, <http://www.tei-c.org/>), jehož cílem je efektivnější zpracování textů. Následuje popis systému Oxfordský textový archiv (Oxford Text Archive, <http://www.ota.ox.ac.uk/>), standardu CES (Corpus Encoding Standard, <http://www.cs.vassar.edu/CES/>) a zmíněny jsou také projekty pro popis rukopisů a historických fondů MASTER (<http://www.tei-c.org.uk/Master/Cataloguing/>) a MEMORIA (http://www.manuscriptorium.com/Site/CZE/default_cze.asp). --

Třetí kapitola, která se věnuje problematice popisných metadatových formátů, představuje formát EAD (Encoding Archival Description, <http://www.loc.gov/ead/>) a s ním související systémy Centrála archivů (Archives Hub, <http://www.archiveshub.ac.uk/>) a Janus (<http://janus.lib.cam.ac.uk/>). -- Závěr shrnuje poznatky o používaných popisných jednotkách a metadatových prvcích, které používají popisované projekty a systémy. Analýza ukázala nejednotnost ve stanovování popisných jednotek i v použití metadatových formátů. Volba popisné jednotky závisí obvykle na cílech systému, potenciálních uživatelích daného systému a jejich předpokládaných potřebách. Většina systémů přejaté formáty přizpůsobila vlastním požadavkům a účelům (přidávání prvků, specifikace prvků pomocí kvalifikátorů, využití souborů autorit a řízených slovníků apod.). Stává se, že pracovníci systémů spolupracují s organizací, která zajišťuje vývoj původního formátu. Podávají připomínky k formátu a návrhy na jeho úpravu. -- Nejednotnost nejen popisných jednotek ale i metadatových prvků a jejich obsahu je obvykle řešena při sdílení a výměně záznamů pomocí konverzních tabulek, které slouží k převodu záznamů z jednoho formátu do formátu jiného. Výrazně se také prosazuje formát pro strukturální metadata METS, který dokáže do jednoho digitálního objektu „sbalit“ všechny soubory, záznamy a informace k jednomu informačnímu zdroji.

Klíčová slova

metadata, metadatové formáty, elektronické informační zdroje, granularita bibliografického popisu, Dublin Core, EAD (Encoding Archival Description), TEI (Text Encoding Initiative), RDF (Resource Description Framework), METS (Metadata Encoding and Transmission Standard)

Abstract

This thesis deals about problems of identification and description of electronic information resources on the Internet. It is mainly focused on web resources. It outlines some specific features of electronic resources and reasons why conventional cataloging methods are not proper for description of these resources. The main goal of this work is to make reader acquainted with selected descriptive metadata formats and illustrate their usage in practice. -- Problems of determination of bibliographic unit are mentioned briefly. There is short chapter on IFLA study Functional Requirements for Bibliographic Records (FRBR, <http://www.ifla.org/VII/s13/wgfrbr/index.htm>). Bibliographic entities and their relations are presented. -- Some technical standards for metadata communication are remembered – mark-up languages, Resource Description Framework (RDF, <http://www.w3.org/RDF/>) and a standard for encoding descriptive, administrative, and structural metadata METS (Metadata Encoding and Transmission Standard, <http://www.loc.gov/standards/mets/>). -- Next chapters turned to selected descriptive metadata formats and following systems and projects. Each format, system or project is briefly described – bibliographic unit, format structure, eventually differences from original format. The chosen metadata formats are mostly used for description of electronic resources at the present time. There are many following systems and projects. Some of them were chosen to illustrate different ways of usage of original formats. -- The first described metadata format is Dublin Core

(<http://www.dublincore.org/>) with following Australian Government Locator Service (AGLS, <http://www.agls.gov.au/>) and Education Network Australia (EdNA, <http://www.edna.edu.au/>). Next format ETD-MS, related also to Dublin Core, is used for Networked Digital Library of Theses and Dissertations (NDLTD, <http://www.ndltd.org/>). Two Czech project WebArchiv (<http://www.webarchiv.cz/>) and ANL FULL (<http://full.nkp.cz/>) are also mentioned within this chapter. -- The second metadata format TEI (Text Encoding Initiative, <http://www.tei-c.org/>) is used for processing whole texts. Work is focused on one part of this format - TEI header - and its application within Oxford Text Archive (OTA, <http://www.ota.ox.ac.uk/>), Corpus Encoding Standard (CES, <http://www.cs.vassar.edu/CES/>) and description for manuscripts and historical collections in projects MASTER (<http://www.tei-c.org.uk/Master/Cataloguing/>) and Czech project MEMORIA (http://www.manuscriptorium.com/Site/CZE/default_cze.asp). -- The last chapter is focused on metadata format EAD (Encoding Archival Description, <http://www.loc.gov/ead/>) and systems Archives Hub (<http://www.archiveshub.ac.uk/>) and Janus (<http://janus.lib.cam.ac.uk/>). -- Conclusion summarises information on used bibliographic units and metadata elements. Differences are at both fields and depend on goals of particular system or project, users and their needs. Original formats are usually adapted (new elements, qualifiers or schemas etc.). It happens that experiences of some project or system become a part of changes during development of original format. -- Disunity of granularity and metadata elements or its content is solved by data conversion between formats and in presence also by format METS. METS can hold together different files, records and other data of one information resource.

Keywords

metadata, metadata formats, electronic information resources, granularity of bibliographic description, Dublin Core, EAD (Encoding Archival Description), TEI (Text Encoding Initiative), RDF (Resource Description Framework), METS (Metadata Encoding and Transmission Standard)

Obsah

PŘEDMLUVA.....	1
1 ÚVOD	4
1.1 PROBLÉMY UCHOVÁVÁNÍ A ZPŘÍSTUPŇOVÁNÍ INFORMACÍ	4
1.2 POPIS ZÁZNAMŮ O ZDROJÍCH.....	5
1.3 VYHLEDÁVÁNÍ V PROSTŘEDÍ WWW	7
1.4 ARCHIVACE A ZPŘÍSTUPNĚNÍ ELEKTRONICKÝCH ZDROJŮ	8
2 POPISNÉ BIBLIOGRAFICKÉ JEDNOTKY	10
2.1 POPISNÉ JEDNOTKY A GRANULARITA	10
2.2 STUDIE FRBR IFLA	10
2.2.1 <i>Objekty studie FRBR - entity</i>	11
2.2.1.1 Vlastnosti bibliografických entit.....	14
2.2.1.2 Vztahy mezi entitami.....	15
2.2.2 <i>Význam studie FRBR</i>	17
3 SOUVISEJÍCÍ TECHNICKÉ STANDARDY	19
3.1 ZNAČKOVACÍ JAZYKY A JEJICH VÝZNAM PRO METADATA	19
3.2 ZNAČKOVACÍ JAZYK SGML.....	20
3.3 ZNAČKOVACÍ JAZYK XML	21
3.4 RDF	24
3.5 PROTOKOL OAI-PMH.....	29
3.6 SCHÉMA METS	31
4 FORMÁT DUBLIN CORE A JEHO APLIKACE.....	35
4.1 FORMÁT DUBLIN CORE	35
4.1.1 <i>Popisné jednotky</i>	39
4.1.2 <i>Metadatové záznamy</i>	39
4.1.3 <i>Shrnutí informací k formátu Dublin Core</i>	43
4.2 FORMÁT AGLS	45
4.2.1 <i>Popisné jednotky</i>	47
4.2.2 <i>Metadatové záznamy</i>	48
4.2.3 <i>Shrnutí informací k formátu AGLS</i>	53
4.3 EDUCATION NETWORK AUSTRALIA (EdNA).....	54
4.3.1 <i>Popisné jednotky</i>	56
4.3.2 <i>Metadatové záznamy</i>	56
4.3.3 <i>Vyhledávání</i>	59
4.3.4 <i>Shrnutí informací k portálu EdNA</i>	61

4.4	POPIS VYSOKOŠKOLSKÝCH PRACÍ NA PŘÍKLADU NDLTD	62
4.4.1	Archivace vysokoškolských prací.....	62
4.4.2	NDLTD.....	63
4.4.2.1	Metadatové záznamy	65
4.4.2.2	Vyhledávání	68
4.4.2.3	Shrnutí informací k systému NDLTD	69
4.5	WEBARCHIV	71
4.5.1	Popisné jednotky	75
4.5.2	Metadatové záznamy	75
4.5.3	Vyhledávání.....	76
4.5.4	Shrnutí informací k projektu WebArchiv.....	77
4.6	PROJEKT ANL FULL.....	79
4.6.1	Popisné jednotky	82
4.6.2	Metadatové záznamy	82
4.6.3	Vyhledávání.....	84
4.6.4	Shrnutí informací k projektu ANL FULL	85
5	HLAVIČKY TEI A JEJICH APLIKACE	86
5.1	FORMÁT TEI.....	86
5.1.1	Popisné jednotky	87
5.1.2	Struktura formátu TEI.....	87
5.1.3	Hlavičky TEI	89
5.1.4	Shrnutí informací k formátu TEI	95
5.2	OXFORDSKÝ TEXTOVÝ ARCHIV	96
5.2.1	Popisné jednotky	97
5.2.2	Metadatové záznamy	97
5.2.3	Vyhledávání.....	99
5.2.4	Shrnutí informací k archivu OTA.....	99
5.3	STANDARD PRO KÓDOVÁNÍ KORPUSŮ.....	100
5.3.1	Popisné jednotky	102
5.3.2	Metadatové záznamy	102
5.3.3	Shrnutí informací k formátu CES.....	106
5.4	POPIS RUKOPISŮ A HISTORICKÝCH FONDŮ	107
5.4.1	Projekt MASTER.....	107
5.4.1.1	Bibliografické jednotky.....	108
5.4.1.2	Metadatové záznamy	108
5.4.2	Projekt MEMORIA.....	113
5.4.2.1	Popisné jednotky	114
5.4.2.2	Metadatové záznamy	114
5.4.2.3	Vyhledávání	116
5.4.3	Shrnutí informací k popisu rukopisů	117

6	HLAVIČKY EAD A JEJICH APLIKACE.....	118
6.1	FORMÁT EAD	118
6.1.1	Popisné jednotky	120
6.1.2	Struktura formátu EAD	120
6.1.3	Hlavičky EAD	122
6.1.4	Shrnutí informací k formátu EAD	124
6.2	CENTRÁLA ARCHIVŮ	125
6.2.1	Popisné jednotky	128
6.2.2	Metadatové záznamy	128
6.2.3	Vyhledávání	132
6.2.4	Shrnutí informací k Centrále archivů	134
6.3	JANUS.....	135
6.3.1	Popisné jednotky	136
6.3.2	Metadatové záznamy	136
6.3.3	Vyhledávání	138
6.3.4	Shrnutí informací k projektu Janus	140
7	ZÁVĚR.....	141
7.1	POPISNÉ JEDNOTKY	141
7.2	METADATOVÉ FORMÁTY	143
	POUŽITÉ ZKRATKY	146
	SEZNAM POUŽITÉ LITERATURY	149
	PŘÍLOHY	

PŘEDMLUVA

Rigorózní práce pojednává o způsobu identifikace a popisu informačních zdrojů, které jsou dostupné prostřednictvím sítě Internet a to především přes službu www. Práce se zaměřuje na popisné metadatové formáty, které slouží pro popis elektronických informačních zdrojů, a na jejich aplikaci v informačních systémech na www. Důraz je kladem především na porovnání použitých metadatových popisů a úpravy použitých formátů oproti jejich základní podobě.

Otázka identifikace a popisu nejen elektronických zdrojů je především s ohledem na následné vyhledání a použití zdrojů velmi aktuální. V zahraničí i v České republice můžeme najít řadu různých projektů a systémů, které se zabývají vytvářením repozitářů s uloženými elektronickými zdroji a databází, které obsahují záznamy o elektronických zdrojích. Hlavním cílem je tyto zdroje vyhledat a využít. Předpokladem této možnosti je samozřejmě dlouhodobé uchování informačních zdrojů a zajištění jejich dostupnosti a čitelnosti.

Informační instituce a knihovny používají pro popis informačních zdrojů různé formáty. V současnosti se rovina popisu informačních zdrojů stále rozšiřuje. Dříve bylo pro knihovníky a informační pracovníky nezbytné orientovat se ve formátech typu MARC, dnes se stává důležitou také orientace v metadatových formátech.

Rigorózní práce vychází z mé diplomové práce (Vozková, 2002) a její původní kořeny je možno hledat ve studijně rozborové práci (Vozková, 2001b) a bibliografické rešerši (Vozková, 2001a), které jsem zpracovala během magisterského studia. Rigorózní práce reaguje na změny v oblasti metadatového popisu. Upustila jsem od analýzy některých formátů a s nimi souvisejících projektů a systémů. Zaměřila jsem se pouze na formáty Dublin Core, EAD a TEI, které jsou v současnosti pro popis elektronických zdrojů nejvíce využívány. Zrevidovala jsem také výběr navazujících projektů a systémů.

Z práce byla vypuštěna analýza britského projektu *Organizace a vyhledávání zdrojů v předmětově založených službách* (Resource Organisation and Discovery in Subject based Services, ROADS). Tento projekt byl ukončen. Byl určen pro usnadnění katalogizace informačních zdrojů na Internetu. Původně na něm spolupracovaly Britský úřad pro knihovní a informační síť (The UK Office for Library and Information Networking, University of Bath, UKOLN), Institut

vzdělávání a výzkumné technologie (The Institute of Learning and Research Technology, University of Bristol, ILRT) a Oddělení pro počítačová studia (The Department of Computer Studies, Loughborough University). Projekt byl zahájen v roce 1995 jako část programu eLib (Electronic Libraries), v jehož rámci byl také do července 1999 sponzorován (Kirriemuir, 1997). Systémy, které používaly tento formát, již většinou přešly na jiné formáty nebo vytvořily formáty vlastní. Spolu s projektem ROADS byla z práce odstraněna kapitola věnovaná britskému portálu SOSIG (Social Science Information Gateway), který formuláře ROADS používal. Portál SOSIG byl sloučen s portálem Altis a dále transformován do portálu Intute, konkrétně jeho části věnované sociálním vědám.

Upustila jsem také od analýzy projektů, které používají metadatový formát CIMI (Consortium for the Computer Interchange of Museum Information). 15. prosince 2003 ukončilo konsorcium CIMI svoji činnost. Hlavním důvodem byl nedostatek financí. Projekty a systémy, které formát CIMI využívaly, se pohybují v oblasti muzejních objektů. Popis muzejních objektů je velmi složitý a poněkud se vymyká oblasti, na kterou se zaměřuje tato práce. Na Ústavu informačních studií a knihovnictví vznikla rigorózní práce, která se na využití informačních technologií v muzejnictví zaměřuje (Šedo, 2004).

Vypustila jsem také kapitolu o projektu Early Canadiana Online (ECO). Databáze ECO sice stále existuje a podle rostoucí návštěvnosti uživatelů je možné soudit, že je úspěšná. Nicméně již používá místo záznamů v metadatovém formátu EAD uživatelské záznamy v XML, které reprezentují záznam MARC a obsahují specifikaci každé obrazové stránky (výšku, šířku, signaturu MD5,...), typ a číslo stránky a informace o textu, které byly z dané obrazové stránky získány pomocí metody OCR. Záznamy jsou určeny především pro interní použití v systému. Přestože je možné záznamy transformovat pomocí XSLT, o jejich sdílení nebo výměnu prý zatím nebyl zájem (Wueppelmann, 2006).

Práce je rozdělena na několik částí. Počáteční kapitoly jsou úvodem do problematiky popisu elektronických zdrojů (kapitola 1), granularity a určení popisné jednotky (kapitola 2) a metadatových formátů obecně (kapitola 3). Krátce shrnují také základní informace o souvisejících technických standardech (kapitola 4). Jde především o stručné nastínění problematiky značkovacích jazyků, které hrají významnou úlohu při ukládání a přenosu metadatových záznamů, dále o rámec pro popis zdrojů RDF, protokol nejen pro sklizení metadatových záznamů OAI-PMH

a o strukturální metadatové schéma METS. Následující kapitoly se věnují nejdříve vybranému metadatovému formátu (Dublin Core, EAD, TEI) a dále projektům a systémům, které popsany metadatový formát používají. Každý formát je krátce představen a analyzován z hlediska popisných jednotek a způsobu jejich popisu. Práce se zaměřuje na použitý metadatový formát. U systémů či projektů jsou navíc uvedeny nejvýznamnější odlišnosti od výchozího formátu. Závěr shrnuje celkové výsledky analýzy popisných jednotek i metadatových formátů.

K projektům a systémům jsou připojeny ukázky záznamů. Kratší ukázky jsou umístěny v textu, aby byly co nejbližší popisu prvků formátu. Obsáhlejší ukázky záznamů jsou uloženy v přílohách. Kromě záznamů obsahují přílohy některé tabulky (např. výčet prvků formátu), případně doplňující obrázky.

Při přípravě této práce jsem čerpala informace především z dostupných elektronických časopisů, domovských webových stránek formátů, systémů a projektů a konferenčních příspěvků. Na webových stránkách věnovaných metadatovým formátům bývají alespoň částečné seznamy s odkazy na projekty a systémy, které daný formát využívají. Vzhledem k povaze popisovaných projektů a systémů byla většina informací čerpána z dokumentace, která je dostupná na domovských webových stránkách těchto projektů a systémů. Některé problémy musely být konzultovány přímo s řešiteli projektů a garanty systémů. Protože větší část popisovaných projektů a systémů je provozována v zahraničí, probíhaly konzultace obvykle formou elektronické pošty. Bohužel, ne vždy se mi podařilo získat všechny potřebné informace. Vzhledem k tomu je rozsah popisu a analýzy popisovaných systémů velmi rozdílný. Některé systémy jsou popsány podrobněji než jiné.

Čerpala jsem většinou z anglicky psané literatury. I v českých textech jsou pro některé termíny používány anglické výrazy. Snažila jsem se většinu z nich přeložit. Protože jsou některé termíny a názvy obtížně přeložitelné do češtiny, je jejich překlad volný.

Použité zdroje jsou zpracovány v souladu s normou ISO 690 a ISO 690-2. V práci jsou odkazovány pomocí prvního údaje záznamu a data vydání (tzv. Harvardský systém).

Za podnětné konzultace k rigorózní práci a laskavý a trpělivý přístup bych chtěla velmi poděkovat PhDr. Evě Bratkové.

1 ÚVOD

1.1 Problémy uchovávání a zpřístupňování informací

Spolu s velkým nárůstem klasických i elektronických dokumentů a zdrojů, které zaznamenávají informace, vývoj a kulturní hodnoty a postoje lidstva, roste také význam dlouhodobého uchovávání a zpřístupňování těchto zdrojů a především informací, které nesou. V některých případech je však kromě informace zaznamenané uvnitř zdroje nepostradatelná také informace o formě, která informaci nese.

Lidé se pokoušeli již od pradávna zachycovat své myšlenky a předávat si informace nejrůznějšími způsoby. Dnešní moderní doba dospěla ke snaze maximum informací z jiných forem převést a uchovat v digitální formě. Domníváme se, že takto zaznamenané informace bude možné uchovávat a zpřístupňovat po staletí, tisíciletí a snad i déle.

Nelze pominout fakt, že možnost šířit informace v digitální formě urychlila pokrok. Vědci a výzkumníci mají možnost rychleji přistupovat k novým vědeckým poznatkům, mohou rychleji komunikovat mezi sebou (někdy dokonce v reálném čase) na velké vzdálenosti, situace dovoluje vznik vědeckých týmů složených z vědců různých národností, kteří spolu komunikují pomocí informací v digitální formě.

Není možné jednoznačně říci, že digitální forma je pro uchovávání informací zcela nejlepší. V případě digitalizace dokumentu se ztrácí některé důležité podněty, které jsou s fyzickou formou spojeny. Digitalizovaný dokument nebo objekt vnímáme pouze zrakem a sluchem. Ostatní vjemy jsou zcela vyřazené, nebo přinejmenším podstatně zkreslené. Navíc jsou s otázkou digitalizace spojeny i značné finanční, časové a technické nároky a náklady. Jednodušší je situace v případě, že dokument nebo objekt jako digitální vznikl.

Aby bylo možné elektronické zdroje vyhledat a zpřístupnit uživateli, musí hledání předcházet vytvoření záznamu o zdroji. Vytvoření může být jak automatické tak ruční. V případě automatického zpracování zdroje jde buď o fulltextové indexování, nebo je výsledný záznam vytvářen na základě určitých charakteristik zdroje. Obvykle se automatické zpracování opírá o předpoklady, že základní údaje

jsou umístěny na určitých místech (název, autor apod.). Pro věcné zpracování zase přicházejí v úvahu kvantitativní metody zjišťující např. četnost slov v textu. Při manuálním popisu by měla technika sloužit především jako pomůcka pro sjednocení postupů, metodik a různých přístupů jednotlivých katalogizátorů. Příkladem mohou být různé rejstříky, tezaury a třídění, které se využívají jak při popisu tak při vyhledávání zdrojů.

Otázka zpřístupnění vyhledaného zdroje souvisí i s autorskými právy a měla by být řešena již v rámci identifikace a popisu zdroje, případně jako navazující krok popisu.

Identifikace a popis zdrojů jsou důležité také v případě výměny nebo sdílení záznamů mezi institucemi ať už na regionální, národní nebo mezinárodní úrovni. Aby byla výměna a sdílení záznamů jednodušší, existují a vznikají různé metodiky, pravidla, standardy a doporučení, která mají vést k jednotnému popisu. Kde není možné popis sjednotit, vstupují do hry různé převodní mechanismy a technické standardy. Standardizace je pro výměnu a sdílení záznamů důležitá. Umožňuje širší spolupráci více institucí, vývoj společných systémů, softwaru, případný vznik konsorcií. Dovoluje tím ušetřit nemalé finanční částky.

Vzhledem k různorodosti institucí a jejich potřeb však není možné na standardizaci přespříliš lpět. Vždy záleží na konkrétní instituci, jaké standardy a normy přijme, nakolik chce a také může spolupracovat s ostatními institucemi.

1.2 Popis záznamů o zdrojích

Spolu s prvními sbírkami dokumentů nebo jiných informačních zdrojů se projevila nutnost tyto sbírky třídit a prvky sbírek popisovat, aby bylo později možné je snadněji vyhledat a zpřístupnit.

Zpočátku byly záznamy velmi stručné a obsahovaly pouze základní údaje o objektech sbírek. Postupem času se tyto záznamy rozšiřovaly o mnoho dalších prvků, které se ukázaly užitečné pro vyhledávání a zpřístupňování.

Aby byla alespoň částečně odstraněna roztržitost popisu a zajištěna obecná srozumitelnost popisu, byla na národní i mezinárodní úrovni vyvíjena doporučení a pravidla popisu. Tato pravidla se stala nezbytnými v době, kdy se začalo s používáním počítačů v knihovnách a vývojem formátů pro bibliografické a katalogizační záznamy.

Významné období ve vývoj popisu zdrojů stále představují 60. léta 20. stol., kdy vznikl v Kongresové knihovně ve Washingtonu výměnný formát MARC (Machine Readable Cataloguing). Znamenají začátek převodu bibliografických záznamů do strojem čitelné podoby. MARC je výměnný formát pro data ve strojové podobě. Stanoví strukturu uspořádání dat, sémantiku a znakovou sadu.

Původní MARC formát se stal základem pro řadu dnes používaných formátů. Kromě této větve existují ještě další typy formátů. Vedle skupiny formátů MARC existují také formáty NON-MARC. S rozvojem Internetu a digitálních technologií se stala velmi aktuální třetí skupina, kterou představují novodobé formáty pro elektronické zdroje, tzv. metadata. Je to skupina, na kterou se zaměřuje tato práce.

Metadatové formáty úzce souvisí s rozmachem Internetu, elektronickým publikováním a budováním digitálních knihoven a archivů. Vznikly, protože popis elektronických zdrojů pomocí stávajících typů formátů nebyl plně vyhovující. Elektronické zdroje mají řadu specifických vlastností. Zmínme alespoň interaktivitu, globální dostupnost, možnost snadného provedení úprav, multimedialitu. Je možné propojovat je pomocí hypertextu mezi sebou i pomocí odkazů propojovat určitá místa uvnitř jednoho dokumentu. Zdroj může sdružovat různé typy vyjádření - text, zvuk, obraz, které mohou být uloženy v různých souborech. Vznikají tak multimediální zdroje. Mezi problémy elektronických zdrojů patří jejich často krátká životnost, dále možnost náhlých změn URL nebo změn verzí. Snadná dostupnost a editovatelnost elektronických zdrojů představují zvýšené riziko plagiátorství. Nabízí zároveň ale také rozšířené možnosti pro vědu, výzkum a vzdělávání. Je však vždy nutné ověřit kvalitu elektronického zdroje, protože publikování prostřednictvím Internetu je velmi snadné. Proto i při výběru zdrojů pro popis je nutné ověřovat důvěryhodnost zdroje a autority, které za obsah zdroje ručí.

Jednotný popis informačních zdrojů nemá zajistit pouze jejich snadnější čitelnost a vyhledatelnost, ale také usnadnit výměnu těchto záznamů mezi institucemi. Zpočátku fungovala výměna záznamů dávkově kopírováním záznamů na externí média a nahráváním záznamů na cílovém pracovišti. Nyní se využívá sítí mezi počítači, především sítě Internet, pro přenášení dat přes protokol Z39.50 nebo protokol OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting). Někdy se používá přímo katalogizace do sdílené báze. Všechny tyto postupy mají urychlit, zjednodušit a sjednotit vytváření a sdílení záznamů. Zároveň

šetří finanční náklady zpracovatelských institucí a usnadňují vyhledání a zpřístupnění informačního zdroje.

1.3 Vyhledávání v prostředí www

Správně vytvořený popisný záznam je nezbytný pro následné vyhledávání dokumentů. Klasické metody prohledávání lístkových katalogů a jiných tištěných nebo ručně psaných vyhledávacích pomůcek nahradily z velké části elektronické katalogy a databáze. Vyhledávání v nich zajišťují různé mechanismy. Od prvních dob, kdy bylo možné používat základní typy dotazů se dostáváme do doby, kdy se počítačová experti pokouší vytvořit sofistikované aplikace, které dovolí nejen klást složité dotazy, ale zároveň také pomohou uživateli s procesem dotazování jako takovým. Do vyhledávacích strojů jsou zabudovávány mechanismy pro automatické opravy chybně zapsaných termínů, navigování a nabízení nových slovních spojení a termínů. Tyto mechanismy jsou založeny na důkladné znalosti lingvistiky. Užívají se také poznatky z jiných oborů – např. matematiky, statistiky atd.

Pro vyhledávání netříděných materiálů v prostoru Internetu a především webu se používají vyhledávací stroje a hierarchicky uspořádané předmětové katalogy. Předmětové katalogy jsou zpracovávány obvykle ručně, což je značně drahý a zdoluhavý proces. Předmětové katalogy nemohou obsáhnout tak velké množství zdrojů jako indexovací a vyhledávací stroje.

Vyhledávací stroje prohledávají dle stanovených kritérií dokumenty na Internetu. Dokáží projít a zaindexovat celý dokument, takže je možné vyhledávat v plném textu zdroje, pokud je tento zdroj textový. Relativně úspěšně jsou prohledávány i obrazové zdroje. Stroje pracují s rozsáhlými automaticky tvořenými seznamy slov a slovních spojení, které jsou propojeny se zdroji. To s sebou nese ovšem riziko sémantického nepochopení, protože jazyk indexace není v takovém případě formalizovaný. Stroje pracují s možnostmi upřesnění dotazu, jako jsou např. frekvence termínu či umístění slova v textu, nicméně ani pomocí těchto omezení dotazu nezabráníme množství nerelevantních zdrojů ve výsledku vyhledávání. Úspěšnější vyhledávací stroje používají pro vyfiltrování relevantních zdrojů metody vážení, statistické metody, sledování počtu odkazů na zdroj apod.

Některé vyhledávací stroje zahrnují do vyhledávání a hodnocení relevance zdrojů také metadatové záznamy, které jsou ve zdrojích obsaženy.

Význam metadatových záznamů je natolik zřejmý, že existují i projekty zaměřené na vytvoření metadatových vyhledávacích strojů. Příkladem může být metadatový vyhledávací stroj MKSearch, jehož beta verze byla zveřejněna koncem roku 2005. Vyhledávač vzniká v rámci stejnojmenného projektu. Jde o open source vyhledávač, který z dokumentů extrahuje metadatové záznamy Dublin Core a ukládá je ve formátu RDF. Tyto záznamy následně používá při vyhledávání (Shaw, 2005).

1.4 Archivace a zpřístupnění elektronických zdrojů

Samotný popis elektronického zdroje pro úspěšné uložení, archivaci a zpřístupnění zdroje nestačí. Je nutné také zajistit jeho stálou dostupnost a čitelnost v nespecifikovaném časovém okamžiku a to s ohledem na autorská práva a na práva uživatele, který se k informacím chce dostat.

Při archivaci elektronických zdrojů se používají metody konzervace, emulace, migrace a zaheslování.

Technologie konzervace spočívá v archivaci zdroje spolu s jeho originálním hardwarovým i softwarovým prostředím. Zdroj je tak opakovaně přístupný a čitelný ve svém původním prostředí a nezměněn. Tato technologie se používá spíše pro krátkodobou archivaci.

Podstatou emulace je vyvinutí programu, jenž napodobuje chování původního hardwaru a softwaru, pro který byl zdroj uzpůsoben. Pomocí série takovýchto programů by mělo být možné, aby byl elektronický zdroj spuštěn na aktuální platformě.

Migrace představuje stálé převody ze zastarávajících technologií na technologie nové. Je to nejvíce testovaná metoda. Jde o aktualizaci formy uložení zdroje tak, aby byla kompatibilní s novým softwarem či hardwarem (např. převod z T602 do Wordu, nebo z nižší verze programu do verze vyšší).

Metoda zaheslování předpokládá, že elektronický objekt je do jisté míry schopen popsat sám sebe. Informační zdroj je zhuštěn a zaheslován do té míry, aby byly zachovány informace, podle kterých může být zdroj opět dekodován (Day, 2004).

O archivaci dokumentů zdaleka nejde pouze v oblasti archivnictví, knihovnictví a muzejnictví. I ve sféře podnikové jsou řešeny otázky převodu papírových dokumentů do elektronické podoby a archivace papírových i elektronických dokumentů. Mezi dokumenty se vytváří vztahy a vazby, jsou evidovány různé verze dokumentů a je nutné zajistit vždy orientaci v těchto vztazích a rychlou dostupnost a dohledání elektronického dokumentu i případného papírového originálu. I pro tuto oblast jsou metadatové záznamy nepostradatelné. Zajišťují nejen správu, archivaci a vyhledatelnost zdrojů, ale často jsou používána také ke kontrole přístupových práv ke zdroji.

2 POPISNÉ BIBLIOGRAFICKÉ JEDNOTKY

2.1 Popisné jednotky a granularita

Budeme-li se zabývat popisem dokumentů a zdrojů, brzy zjistíme, že určit popisnou jednotku není vždy zcela snadné. Musíme vycházet z nosného média, obsahového celku jednotky a zahrnout potřeby uživatelů. Mezi potřeby uživatelů patří *nalezení* relevantních informací korespondujících s jejich problémem, *identifikace* konkrétního zdroje odpovídajícího vyhledanému záznamu, *výběr* zdroje, který je přiměřený možnostem a schopnostem uživatelů, nebo *získání* daného zdroje. Uživatelské potřeby se mohou vzájemně doplňovat (Mezinárodní federace knihovnických asociací a institucí, 1998).

Při popisu zdrojů narážíme na problém tzv. granularity (zrnitosti). Každý katalogizátor řeší otázku, na jak detailní úrovni zdroj popisovat. Jde nejen o problém, zda popisovat jednotlivé elektronické zdroje nebo celý systém propojených elektronických zdrojů se stejnou tematikou, příp. jednotlivý obrázek. Jde také o problém, jak popisovat např. reprodukci díla, kopie díla apod. Zajímá koncového uživatele to, že se jedná např. o obraz *Slunečnice* od Van Gogha, nebo je pro něho důležité, kdo a kdy provedl digitalizaci tohoto obrazu? Tyto problémy může řešit každá organizace jiným způsobem. Rozdíly v záznamech a jejich pojetí působí problémy při spolupráci a výměně záznamů.

2.2 Studie FRBR IFLA

Mezinárodní federace knihovnických asociací a institucí IFLA se zabývá bibliografickým popisem a mezinárodní bibliografickou kontrolou již od roku 1961, kdy byla v rámci konference IFLA v Paříži stanovena první katalogizační doporučení známá dnes jako *Pařížské principy* (Paris Principles). Z jiné konference vzešla opět na poli organizace IFLA norma *Mezinárodní standardní bibliografický popis pro monografie* (Mezinárodní federace knihovnických asociací a institucí, 1993). Není divu, že se tato organizace zasloužila i o vývoj a výzkum na poli vztahů informačních entit. Ustanovená skupina IFLA se zabývala od roku 1991 modelem informačních entit, jejich vlastnostmi a vztahy. Navrhla model, ve kterém jsou

zachyceny vztahy mezi entitami, které mohou souviset s popisem nějakého objektu. Model může napomoci vytváření bibliografického záznamu s ohledem na různé uživatelské potřeby a požadavky, na odlišnosti různých typů dokumentů i funkce záznamů. Zároveň zlepšuje díky standardizaci vztahů mezi entitami podmínky pro sdílenou katalogizaci.

Od zveřejnění výsledků skupiny IFLA ve studii *Funkční požadavky na bibliografické záznamy* (Mezinárodní federace knihovnických asociací a institucí, 1998) se vztahy mezi entitami zabývá i řada dalších studií, projektů a konferencí. Ze studie FRBR vyplývají určitá doporučení pro tvorbu bibliografických záznamů národních bibliografií. Národní knihovna České republiky vydala překlad této studie v roce 2002 (Mezinárodní federace knihovnických asociací a institucí, 2002).

V roce 2003 byl na setkání ve Frankfurtu v Německu schválen organizací IFLA návrh ustanovení mezinárodních zásad katalogizace. Toto ustanovení reflektuje zvýšený význam standardizace mezinárodní katalogizace. Její význam dále roste v souvislosti s rozsáhlejšími možnostmi přístupu ke katalogizačním záznamům prostřednictvím OPAC katalogů. Ustanovení představuje vlastně adaptaci původních Pařížských principů z roku 1961. Je rozšířeno o poznatky z modelu FRBR a obdobného modelu pro záznamy o jmenných autoritních entitách FRANAR (Functional Requirements and Numbering for Authority Records). Pokrývá tedy popis bibliografických i autoritních záznamů, vztahy mezi záznamy entit, funkce záznamů v katalogu, zabývá se stručně i zásadami vyhledávání. Záznamy o předmětových autoritních entitách bude ošetřovat model FRSAR (Functional Requirements of Subject Authority Record). Modelem FRSAR se zabývá samostatná pracovní skupina, která byla ustanovena v dubnu 2005 (Mezinárodní federace knihovnických asociací a institucí, 2005).

2.2.1 Objekty studie FRBR - entity

Model FRBR se zabývá informačními objekty, jejich vlastnostmi a vztahy. FRBR zohledňuje různé typy bibliografických a autoritních dat (popisné údaje, jmenné i názvové záhlaví, předmětové i lokační údaje). Orientuje se především na uživatele a jeho potřeby vyhledat, identifikovat, vybrat a zajistit si informační zdroj. Důsledně rozlišuje úroveň vnímání informační entity (dílo, vyjádření, zhmotnění, jednotka) (Byrum, 2000). Model FRBR může být aplikován přímo

na data nebo na jejich zobrazení, což je více doporučováno. Záznam je pak vizuálně hierarchicky rozdělen do úrovní, ve kterých si uživatel postupně vybírá od nejvyšší úrovně až ke konkrétní jednotce.

Informační entity jsou v modelu FRBR rozděleny na tři základní skupiny. V první skupině jsou entity, které vznikají jako výsledek intelektuální nebo umělecké činnosti. Jde o entity *dílo* (work), *vyjádření* (expression), *zhmotnění* (manifestation) a *jednotka* (item).

Druhou skupinu tvoří entity *fyzická osoba* (person) a *korporace* (corporate bodies). Jde o entity zodpovědné za intelektuální nebo umělecký obsah a za šíření a distribuci entit z první skupiny. Mohou to být autoři díla, ti, kdo dílo realizují, producenti fyzických nosičů nebo vlastníci konkrétních jednotek.

Třetí skupina obsahuje entity, které stojí v pozici předmětového vyjádření intelektuálního nebo uměleckého obsahu díla. Do této skupiny patří entity *předmět* (object), *pojmem* (concept), *událost* (event) a *místo* (place) (Bratková, 2001; Drobníková, 2002).

Pro zjednodušení budou entity první skupiny v této práci označovány jako bibliografické, druhé skupiny jako jmenné autoritní entity a entity třetí skupiny jako předmětové autoritní.

Model FRBR se zabývá především bibliografickými entitami. Jmenné autoritní entity a entity předmětové autoritní jsou zmíněny okrajově v souvislostech bibliografických entit. Jak již bylo uvedeno výše, druhou skupinou jmenných autoritních entit se zabývá FRANAR a na modelu pro třetí skupinu předmětových autoritních entit pracuje skupina FRSAR.

V modelu FRBR je *dílo* pojato abstraktně jako intelektuální či umělecký výtvor. Za *dílo* nemůžeme považovat žádný konkrétní fyzický objekt. *Dílo* jako myšlenku vnímáme prostřednictvím jeho různých *vyjádření*. Tato *vyjádření* stejného *díla* mohou být např. malé úpravy díla, doslovné překlady, hudební přepisy apod. Pokud byla do stávajícího *díla* vnesena velká intelektuální činnost, můžeme vzniklé *dílo* považovat za nové. Jde např. o převyprávění díla, adaptace díla pro děti, parodie, hudební variace apod. Definování *díla* nám umožňuje prostřednictvím odkazů na *dílo* propojit jeho různá *vyjádření*.

Vyjádření je uměleckou či intelektuální realizací *díla*. *Vyjádření* může být alfanumerické, zvukové, obrazové, v hudební notaci apod. Změna formy *vyjádření* či jeho obsahu vede ke vzniku nového *vyjádření*. Díky *vyjádření* můžeme odlišit

různá *vyjádření* stejného *díla*, nebo naopak propojit dvě různá *vyjádření* se stejným obsahem.

Fyzickým ztělesněním *vyjádření* určitého *díla* je *zhmotnění*. *Zhmotnění* reprezentuje všechny fyzické objekty jednoho *vyjádření*. Jednomu *vyjádření* může odpovídat více *zhmotnění*, pokud bylo např. použito pro zachycení *vyjádření* více typů médií, nebo bylo písmo *vyjádřeno* v jiné grafické podobě apod. Pomocí *zhmotnění* můžeme popsat vlastnosti souboru *jednotek* i jedinečných *jednotek* (rukopis, originál obrazu) a také znázornit vztahy mezi jednotlivými *zhmotněními* stejného *vyjádření* či *díla*.

Jednotka je konkrétní exemplář určitého *zhmotnění*. Pomocí *jednotky* popisujeme jedinečné vlastnosti jedné konkrétní kopie a její vztahy k ostatním kopiím *díla*.

V modelu FRBR jsou zmíněny i entity autoritní a především jejich vztah k bibliografickým bibliografickým entitám. Autoritní entity mohou být jmenné a předmětové.

Jmenné autoritní entity tvoří *fyzická osoba* (person) a *korporace* (corporate body). Jsou s bibliografickými entitami spojeny vznikem a realizací *díla* zcela jednoznačným způsobem nezávisle na podobě autoritních entit v různých *vyjádřeních*, *provedeních* a *jednotkách* *díla* (Mezinárodní federace knihovnických asociací a institucí, 1998).

Entita *osoba* představuje konkrétní osobu. Entita *korporace* sdružuje osoby spojené s určitou korporací, které za organizaci vystupují a jednají jako jednotka. Kromě klasických korporací a institucí jde také o různé konference, expedice, festivaly, ale také stát, region nebo jinou podobnou jednotku.

Každá jmenná autoritní entita má určité vlastnosti. U *osoby* jde o celou řadu vlastností, z nichž lze pro příklad uvést např. životní data, datum narození, pohlaví, jazyk nebo obor působnosti. Mezi vlastnosti entity *korporace* mohou patřit asociovaná místa nebo data, adresa nebo historie korporace. Stejně tak, jako existují vztahy mezi bibliografickými entitami, existují i vztahy mezi jmennými autoritními entitami navzájem (Bratková, 2001).

Vztahy mezi jmennými autoritními entitami slouží k jednoznačné identifikaci autoritní entity. Pomáhají identifikovat a sdružit všechna *díla* související s danou autoritou. Podrobnější informace o typech vazeb mezi těmito entitami jsou dostupné

v návrhu IFLA *Funkční požadavky na autoritní záznamy* (Mezinárodní federace knihovnických asociací a institucí, 2007b).

Autoritní předmětové entity tvoří *předmět, pojem, událost a místo*. *Předmět* vyjadřuje množinu konkrétních hmatatelných věcí či objektů, kterými se *dílo* zabývá. Entita *pojem* zahrnuje množinu abstraktních myšlenek či vyjádření obsahu *díla*. *Událost* označuje akce a časová určení, jež jsou obsahem *díla*. *Místo* označuje lokaci, kterou se *dílo* zabývá (Mezinárodní federace knihovnických asociací a institucí, 1998).

2.2.1.1 Vlastnosti bibliografických entit

Každá z výše uvedených entit může nabývat různých vlastností. Vlastnosti mohou být vnitřní a vnější. Vnitřní jsou implicitně spojeny s danou entitou a vycházejí z ní samotné, z jejího pojetí a obsahu (např. název). Vnější znaky jsou entitě přidělené na základě vnějších souvislostí či vnějších systémů, které lidé stanovili (např. tematická skupina či deskriptor tezauru). Vlastnosti mohou nabývat více než jedné hodnoty. Mohou se také časem měnit. Je-li vlastnost zaznamenána v záznamu o entitě, lze ji použít pro vyhledávání.

Pro každou informační entitu existují jednoznačné identifikátory nebo je plánována jejich specifikace. Jde např. o plánované a zaváděné použití kódů ISWC¹ pro hudební *díla*, ISAN pro audiovizuální *díla*, ISRC pro zvuková *díla*, DOI nebo Handle pro digitální objekty. Pro *zhmotnění* slouží identifikátory ISBN, ISMN, ISSN či ISRN. Mezi identifikátory pro *jednotku* můžeme zařadit signaturu, přírůstkové číslo či čárový kód. Také jmenné autoritní entity mají své identifikátory. Jde např. o připravované číslo ISADN (International Standard Authority Data Number). Pro standardizovaný zápis předmětových autoritních entit se používají různé tezaury, hesláře a systémy přidělení abecedních či alfanumerických kódů.

Některé vlastnosti a znaky jsou společné pro všechny bibliografické entity, jiné se váží ke konkrétní entitě. Mezi vlastnosti společné několika entitám patří např. názvy či podnázvy, data tvorby, jazyk obsahu. Vlastnosti vážící se k určité entitě jsou např. uživatelské určení, způsob vyjádření obsahu, fyzické vlastnosti

¹ Standard ISWC byl schválen v roce 2001 jako mezinárodní standard ISO 15707 Informace a dokumentace - Mezinárodní standardní kód pro hudební díla (Information and documentation - International Standard Musical Work Code, ISWC). Tvar ISWC má podobu např. T-345246800-1. V současnosti probíhá zavádění kódu ISWC.

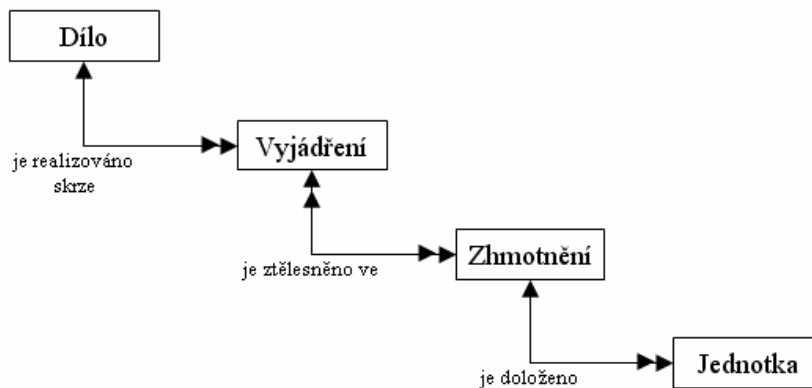
(např. rozsah, rozměry, typ písma či barva). Specifické údaje vážící se ke konkrétní jednotce mohou být např. vlastnictví či omezení přístupu k jednotce (Bratková, 2001).

2.2.1.2 Vztahy mezi entitami

V běžném životě nejsou ustáleny formulace vztahů mezi entitami. Občas nalezneme v jednom zdroji upozornění na vztah k jinému zdroji. V rámci FRBR by měly vztahy sloužit pro propojení a navigaci mezi zdroji a měly by zkvalitnit výsledky vyhledávání. Měly by odkázat uživatele z jednoho zdroje na zdroj související. Vztahy mohou být vyjádřeny pomocí záhlaví, víceúrovňového bibliografického popisu či pomocí poznámek. FRBR požaduje k zajištění platného vztahu jednoznačné určení vztahu obou entit.

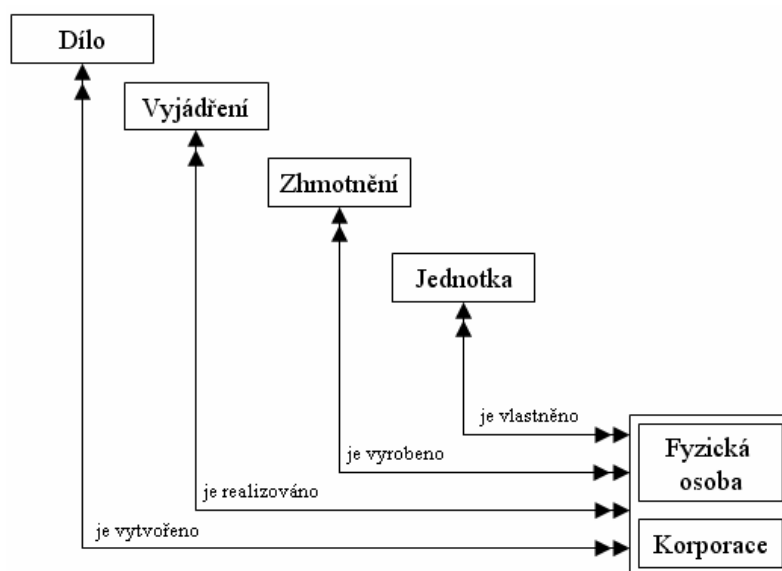
Vztahy mohou být uskutečněny na několika úrovních. Nejvyšší úroveň stanovuje vztahy mezi jednotlivými informačními entitami. Vztahy mezi bibliografickými entitami nejlépe vyjádří jednoduché schéma na obrázku 1 (Mezinárodní federace knihovnických asociací a institucí, 1998; Bratková, 2001).

Obr. 1: Vztahy bibliografických entit dle modelu FRBR



Obrázek 1 znázorňuje, že *dílo* může být uskutečněno pomocí více *vyjádření*, ale *vyjádření* se může vztahovat pouze k jednomu *dílu*. *Vyjádření* je ztělesněno ve *zhmotnění*, kterých může být více. *Zhmotnění* může ztvárňovat více *vyjádření*. *Zhmotnění* je znázorněno pomocí jedné či více *jednotek*. Naopak *jednotka* se váže pouze na jedno *zhmotnění*.

Obr. 2: Vztah bibliografických entit k jmenným autoritním entitám dle modelu FRBR



Od jmenných autoritních entit směřuje ke každé bibliografické entitě jeden typ vztahu. Vztah *vytvořeno* označuje intelektuální či uměleckou odpovědnost za obsah *díla*. Vztah *realizováno* vypovídá o odpovědnosti za provedení *vyjádření*. Vztah *vyrobeno* dokládá odpovědnost za fyzické provedení - výrobu či distribuci, tedy *zhmotnění*. Poslední vztah *vlastněno* vypovídá o vlastnictví či správě *jednotky*. Obrázek 2 ukazuje, že všechny vztahy v tomto případě mohou být vícečetné (Mezinárodní federace knihovnických asociací a institucí, 1998; Bratková, 2001).

Předmětové autoritní entity jsou *předmět*, *pojem*, *událost* a *místo*. Stojí v pozici předmětového vyjádření obsahu *díla*. Všechny informační entity (včetně *díla* samotného) se váží k *dílu* pomocí vztahu *jehož předmětem je*, který vyjadřuje předmětový obsah *díla*. Všechny vztahy mohou být vícečetné.

Mezi informačními entitami existují další vztahy. Jde o vztahy mezi entitami různého typu nebo entitami stejného typu. U vztahů typu *dílo-dílo* vycházíme z toho, že obsah jednoho díla je intelektuálně či umělecky tak odlišný, že je nové *dílo* možno považovat za samostatné. Podle míry novosti můžeme *díla* rozlišovat na samostatná s velkým přínosem a nesamostatná s minimálním odchýlením od *díla* původního. Pro samostatná i nesamostatná *díla* může vztah nabývat hodnot např. *má pokračování/je následující dílo* či *má supplement/je supplementem*. Specifickými hodnotami pro nesamostatná *díla* jsou např. *má imitaci/je imitací*. Tímto vztahem může být vyjádřena i část *díla*. Část *díla* může být závislá

na kontextu celého *díla*, nebo může být část *díla* nezávislá a stát se sama předmětem identifikace a popisu.

Vztah *vyjádření-vyjádření* může reprezentovat různá *vyjádření* stejného *díla* (*má překlad/je překladem*), nebo může jít o *vyjádření* reprezentující odlišná *díla* (*je adaptací/má adaptaci*). I u tohoto vztahu lze brát v úvahu vztahy *celek-část vyjádření* (*má část/je částí*).

Vztah *zhmotnění-zhmotnění* zachycuje *zhmotnění* stejného *vyjádření*. Existují zde vazby *má reprodukci/je reprodukcí* a *má alternaci/je alternací*. Opět nelze zapomenout na možnost vyjádřit vztah *celek-část zhmotnění*, jímž lze např. popsat tištěnou knihu s doprovodným CD-ROMem.

Konečně vztah *jednotka-jednotka* vzniká získáním nové *jednotky* na základě *jednotky* jiné. Jde např. o svázání jednotlivých čísel časopisu či přívazek dvou či více děl. Vazba může být definována jako *má rekonfiguraci/je rekonfigurací*. Jako část *jednotky* je možné popsat např. její kapitolu, obal nebo jinou nesamostatnou část *jednotky*.

Ještě krátce upozorníme na doplňující vztahy. Jde o typ *vyjádření jednoho díla-jiné dílo*, jímž lze určit např. pokračování díla, dodatek, souhrn apod. Vztah *zhmotnění-jednotka* je určen pro případy, kdy reprodukcí *jednotky* vznikne nové *zhmotnění* (Mezinárodní federace knihovnických asociací a institucí, 1998).

2.2.2 Význam studie FRBR

Studiemi FRBR a FRANAR nastínila IFLA široké spektrum možností pro další výzkum. Vliv modelů se začíná projevovat v nových verzích pravidel AACR2 a ISBD, formátech typu MARC a v neposlední řadě také v národních katalogizačních pravidlech. Důležitá je zejména tendence zřetelněji vyjádřit vzájemné hierarchické vztahy mezi entitami. Ozývají se hlasy požadující odklon od popisu reprezentujícího konkrétní *jednotku* k popisu reprezentujícímu soubor *jednotek* na úrovni *vyjádření* či *díla*. Tím by se usnadnila sdílená katalogizace a přebírání záznamů. Spojená řídicí komise pro revizi pravidel AACR připravuje na základě modelu FRBR zcela nová pravidla RDA (Resource Description and Access), která budou zároveň odvozena z pravidel AACR. Vydání standardu RDA je plánované na začátek roku 2009 (RDA, 2007).

O model FRBR projevily zájem i společnosti zajišťující vývoj metadatových formátů a některých identifikátorů. Předpokládá se, že by model FRBR mohl sloužit jako výchozí bod pro vznik komplexnějších modelů, které by obsáhly velké množství dat, podporovaly vyhledávací nástroje a zlepšovaly procesy managementu informačních zdrojů (Byrum, 2000).

Jedním z prvních knihovních systémů, který implementoval model FRBR, byl systém Virtua firmy VTLIS. Bibliografický záznam v něm sestává ze tří částí - *dílo*, *vyjádření*, *zhmotnění*. Může být doplněno holdingovým záznamem - *jednotkou* (VTLIS, 2004). Vyhledávací službou, která podporovala FRBR, byl rešeršní systém RedLightGreen, jehož vývoj byl sponzorován Mellonovou nadací. Zpřístupňoval přes 120 milionů záznamů skupiny RLG (Research Library Group) především pro studenty (Proffitt, 2004). Služba byla ukončena v listopadu 2006 s doporučením, aby její uživatelé dále využívali k vyhledávání WorldCat, který má řadu obdobných rysů (včetně podpory FRBR) jako RedLightGreen.

Činnost skupiny FRBR publikováním studie neskončila. Skupina byla transformována na skupinu FRBR Review Group, která zajišťuje další začleňování studie FRBR do praxe. Kromě vývoje modelu FRBR (detailní provedení modelů pro entity) pomáhá při začleňování terminologie FRBR do katalogizačních pravidel a formátů, spolupracuje s knihovnickými organizacemi a skupinami a vydává doporučení (Mezinárodní federace knihovnických asociací a institucí, 2007a). Jedním z konkrétních výsledků spolupráce se skupinou ISBD je přiřazení prvků ISBD a entit FRBR, které bylo zveřejněno v roce 2004 a je dostupné na stránkách IFLA. Jde o přiřazení jednotlivých popisných oblastí bibliografickým entitám (Mezinárodní federace knihovnických asociací a institucí, 2004).

3 SOUVISEJÍCÍ TECHNICKÉ STANDARDY

3.1 Značkovací jazyky a jejich význam pro metadata

Popisné metadatové formáty slouží k zachycení sémantiky. Určují význam jednotlivých prvků, to, co chceme o zdroji sdělit. Aby bylo možné tuto sémantiku nějak zachytit, vyjádřit a formalizovaně zapsat, používají se značkovací jazyky. Slouží pro vyjádření syntaxe. Třetí důležitou rovinou je vyjádření struktury - vztahů mezi jednotlivými prvky a také mezi různými schématy. Pro vyjádření a ke komunikaci tohoto typu byl vyvinut rámec RDF (Resource Description Framework, <http://www.w3.org/RDF/>).

Značkovací jazyky jsou sady značkovacích pravidel, které se používají pro kódování textu. Jejich počátky můžeme vidět ve značkách, které byly původně užívány k popisu krátkých poznámek a komentářů vázících se k určitému textu. Určovaly, jakým způsobem se má s udaným textem pracovat. Postupně se funkce těchto značek či kódů rozšířily. Obecně můžeme jako kódování textu chápat i použití běžné interpunkce, mezer mezi slovy apod.

Rozlišujeme značkovací jazyky popisné a procedurální. Popisné jazyky rozdělují zdroj na části tím, že jim přiřazují jména. Identifikují část zdroje, např. odstavec. Procedurální jazyky určují, které procesy se mají vykonat na určitém místě zdroje, např. zavolat danou proceduru.

Pro uložení a komunikaci metadatových záznamů se používá formát XML (eXtended Markup Language, <http://www.w3.org/XML/>), který vznikl z jazyka SGML (Standard Generalized Markup Language, <http://www.w3.org/MarkUp/SGML/>). Oba patří mezi skupinu popisných značkovacích jazyků.

Popisný metadatový záznam, který je uložený ve struktuře značkovacího jazyka, je vhodný ke sdílení a výměně mezi různými systémy. Pro přenos takových záznamů mezi servery a databázemi se používají protokoly. V knihovnickém světě se pro přenos záznamů používají především protokoly Z39.50 a protokol OAI-PMH (Open Archive Initiative Protokol for Metadata Harvesting, <http://www.openarchives.org/OAI/openarchivesprotocol.html>), který se pro metadatové záznamy využívá více.

3.2 Značkovací jazyk SGML

SGML (Standard Generalized Markup Language) je mezinárodním standardem pro popis elektronického textu. Je nezávislý na formě reprezentace textu, na hardwaru a softwaru. SGML je obecný soubor pravidel pro vyjádření logické struktury souborů tak, aby bylo usnadněno vyhledávání, výběr a strukturované zobrazení zdroje. Pravidla jsou vyjádřena pomocí značek (tagů) umístěných ve zdroji a vyjadřujících vztahy mezi částmi textu i vazby externí mimo zdroj (EAD, 2002b).

SGML používá jednoduché značky pro označení a identifikaci struktury jednotek textu a přiřazuje jim určitá pravidla používání. Textová jednotka se označuje *prvek* (element). SGML nespecifikuje jména prvků. Prvek je ohraničen *značkami* (tagy). Prvek má obvykle značku uvozující (start-tag) a značku ukončující (end-tag). Značky jsou uzavřeny do lomených závorek a před koncovou značkou je uvedeno lomítko, např. <odstavec>...</odstavec>. Existují ale prvky, které nemají značky párové.

Pro soubory SGML je možné vytvářet definice typu dokumentu, zkráceně DTD (Document Type Definition). Definice typu dokumentu je množina pravidel, která společně definují značkovací jazyk, jeho prvky a způsob uspořádání značek v platném (správném) zdroji (Národní knihovna, 2000). Pomocí DTD by měla být vyjádřena jasná pravidla struktury zdroje pro jeho následnou analýzu a výklad. Stanoví jména prvků, nutnost počátečních a ukončujících tagů, obsahové modely prvků a možnosti četnosti výskytu prvku. Definuje také vlastnosti prvků a jejich případné předdefinované hodnoty.

SGML soubor se skládá ze SGML úvodu a samotného textu zdroje. Úvod obsahuje deklaraci SGML a definici typu dokumentu. SGML popis bývá zakomponován do zdroje, nebo může být umístěn ve zvláštním souboru. Ve druhém případě bývá ve zdroji odkaz na externí DTD. Deklarace SGML specifikuje základní údaje týkající se terminologie znakové sady, kódů oddělovačů, délky identifikátorů apod. (Spergerg-McQueen, 1998).

Jazyk SGML jako obecný značkovací jazyk měl být původně jediným metajazykem. Je však natolik obecný a má tak široký záběr, že není možné pracovat s ním bez mnoha složitých nástrojů. Proto se z tohoto jazyka vyvinuly jiné jazyky, které se omezily pouze na určitou oblast.

Asi nejvíce používanou podskupinou jazyka SGML je HTML. Jazyk HTML vznikl z nutnosti spojit text, grafiku a jiný digitální obsah do jednoho elektronického zdroje a umožnit poslat tento zdroj prostřednictvím sítě Internet. HTML definuje určité standardy, ale přesto je tvorba HTML zdrojů volná a chybí jí stanovení mnoha formálních pravidel. Proto W3C navrhlo nový jazyk XML (eXtensible Markup Language).

3.3 Značkovací jazyk XML

Jazyk XML je rozšiřitelný formální značkovací jazyk založený na jednodušších prvcích jazyka SGML. Předpokládá se, že v budoucnu dojde k přechodu z HTML na XML. Dnes se prozatím pomocí jazyka XML definuje verze jazyka HTML, známá jako XHTML. Jazyk XML umožňuje kontextové vyhledávání. V kombinaci s metadaty je XML mocným nástrojem (Musciano, 2000).

Na rozdíl od jazyka HTML, který má pevně stanovené značky (tagy) v DTD, umožňuje jazyk XML nadefinovat vlastní značky, jimiž lze určit strukturu i význam dat ve zdroji. Způsob zobrazení dat (velikost, barva, zarovnání apod.) je nutné definovat v externím souboru, tzn. definice stylu. Nadefinovaný styl lze použít pro více zdrojů (souborů). Když je změněn styl, změní se i zobrazení ve všech zdrojích, které příslušný styl používají. Možnost oddělení definice stylu zobrazení od významu dat je důležité především pro vyhledávání.

XML má přísnější pravidla než HTML. Pokud zdroj splňuje pravidla XML, označuje se jako správně strukturovaný (well-formed). V XML musí být všechny prvky párové. Výjimkou jsou pouze prázdné prvky bez obsahu, jež jsou však ukončeny znakem “/>“. Důležité je také to, aby se nekřížily jednotlivé tagy. Dále je v XML nutné rozlišovat malá a velká písmena (Kosek, 1999).

Zdroj XML se skládá z deklarací, prvků a jejich atributů, komentářů, instrukcí pro zpracování jinou aplikací a vlastního textu (Kosek, 1999). Pro určení, jaké prvky, vlastnosti prvků a jejich kombinace lze použít, se užívají definice typu dokumentu (Document Type Definition, DTD). DTD definuje čtyři typy deklarací. Jde o deklarace entit, prvků, atributů a notací.

Entita je tvořena jménem a jeho hodnotou. Entita v sobě může nést znakový řetězec, který je přiřazen jinému určenému slovu či řetězci. V tom případě se kdekoli v textu za definované jméno dosadí specifikovaný řetězec znaků z entity

(např. v HTML `<` se nahradí znakem `<`). Existují dva typy entit - analyzované, které procesor zpracovává, a neanalyzované, které procesor ignoruje. Analyzované entity mohou být obecné a parametrické. Obecné entity se užívají přímo ve zdrojích XML (např. `©` pro grafický znak copyrightu). Parametrické se užívají v definicích DTD. Před jejich jménem se uvádí znak `%`. Tento znak znamená, že procesor má najít jméno entity v seznamu parametrických entit dané definice DTD, vložit hodnotu entity místo odkazu na entitu a výslednou hodnotu entity zpracovat jako součást DTD. Příklad 1 obsahuje ukázkou parametrické entity. Každý výskyt entity `&ovoce;` bude v pomyslném následujícím textu nahrazen frází *granátové jablko* a kdekoli se v definici DTD objeví odkaz na entitu `%TypObsahu;` nahradí se slovem *CDATA*.

Příklad 1: Příklad parametrické entity

```
<!ENTITY ovoce "granátové jablko">
<!ENTITY % TypObsahu "CDATA!">
```

Další částí deklarovanou v DTD jsou prvky. Jsou to definice značek. DTD uvádí jméno značky, typ jejího obsahu a gramatiku daného prvku. Gramatika prvků stanoví pravidla opakování, povinnost výskytu prvku, pořadí prvků, výběr platných prvků z určité skupiny a jejich seskupování.

Prvky mohou mít přiřazeny atributy. Ty ovlivňují vzhled a chování prvku. Atribut se skládá ze jména a jeho hodnoty. Pro definici atributů v XML se používá značka `<ATTLIST>`. Existuje několik typů atributů:

- CDATA (libovolný znak nebo řetězec znaků)
 - ID (identifikátor jedinečný v rámci zdroje)
 - IDREF nebo IDREFS (odkaz na seznam identifikátorů, resp. na jiné části zdroje)
 - ENTITY (hodnota definovaná jako entita jinde ve zdroji, pro odkazy na část textu, část deklarace DTD nebo externí soubor)
 - NMTOKEN nebo NMTOKENS (platné jméno XML nebo seznam jmen)
- (Musciano, 2000)

DTD je možné přizpůsobit požadavkům různých organizací a skupin a vytvořit tak pro ně vlastní podskupiny jazyka XML. Jazyk XML je tedy možné chápat i jako metajazyk, jehož pomocí je možné definovat pravidla a prvky pro jiné značkovací jazyky, které jsou přizpůsobeny požadovaným účelům. Definice DTD popisuje nový značkovací jazyk na bázi XML.

Jedno ze základních pravidel DTD pro XML zní, že každá značka (tag) musí mít svoji koncovou značku. Aby byl zdroj platný, musí obsahovat pouze značky, které jsou definovány v deklaraci DTD nebo v deklaraci přímo ve zdroji XML. Na začátku zdroje musí být uvedena použitá verze jazyka XML a značka popisující umístění odpovídající DTD.

Hlavní přednosti DTD v jazyce XML spočívají v možnosti přizpůsobit si DTD a vytvořit si jazyk, který bude odpovídat potřebám oborově zaměřené skupiny. DTD má však několik nedostatků. Např. nedovoluje rozlišit celá a desetinná čísla, určit délku řetězce, nepodporuje některé znakové sady a nové technologie XML jako jmenné prostory (viz níže). Proto se dnes začínají místo DTD používat XML schémata (také XSD). Je to nový jazyk, který popisuje obsah a strukturu XML zdrojů. Všechny stávající XML zdroje vytvořené pomocí DTD je možné převést do XML schémat. XML schémata v sobě mají zabudované některé základní datové typy. Tyto typy vycházejí z datových typů používaných v jazyku SQL a programovacím jazyku Java. Mezi základní typy patří např. řetězec (string), reálná čísla (real), celá čísla (integer), datum/čas (dateTime), URI (uri), jazyk (language). Uživatel může nadefinovat další datové typy. XML schémata dovolují definovat počet výskytů určitého prvku (např. chceme-li stanovit maximální počet prvku *autor* na tři výskyty), podporují import a export externích schémat XML², dovolují v jednom zdroji kombinovat více schémat XML a pracují s objektově orientovanými koncepty (např. s dědičností znaků).

Protože jména prvků a jejich atributů mohou být v XML tvořena volně, je jejich identifikace zajištěna tzv. jmennými prostory. Jmenné prostory dovolují v jednom zdroji kombinovat více sad prvků, čehož využívá mnoho aplikací XML (např. RDF). Každá sada prvků musí mít své jednoznačné URI. Sady jsou definovány obvykle na začátku zdroje pomocí URI a jednoznačného prefixu. Pro deklaraci jmenných prostorů jsou rezervované atributy začínající *xmlns*. Každému prvku dané sady je v popisu předřazen prefix sady, ke které patří.

Pro popis vzhledu a definice dat při přenosu ve formátu XML se používají definice stylů pomocí jazyků CCS (Cascading Style Sheets - kaskádové styly), XSL (eXtensible Stylesheet Language - rozšiřitelný jazyk šablon stylů) a DSSSL (Document Style Semantics and Specification Language). CSS jsou běžné kaskádové

² Datové typy, prvky, atributy a entity, které jsou definované v externích schématech

styly užívané v HTML. DSSSL byl definován pro SGML, je však použitelný i s XML. Je to velmi rozsáhlý stylový jazyk. Dnes se stále častěji užívá jazyk XSL, který je jednodušší než DSSSL. Navíc je možné zapisovat ho přímo v jazyce XML (Kosek, 1999).

Pro kontrolu správné syntaxe XML zdroje slouží nástroj nazvaný parser. Kontroluje syntaxi na základě DTD nebo schématu XML. Parsery mohou existovat jako samostatné programy, ale obvykle bývají zabudovány v prohlížečích, které XML podporují.

XML zdroje je možné konvertovat do jiných formátů, např. HTML, PDF atd. Pro účely konverze slouží procesory XSL. Ke konverzi využívají transformační část XSL označovanou jako XSLT. XSL a XSLT tedy definují jak se text zobrazí (např. barva, velikost), ale dokáží také vypočítat součty čísel, změnit formát data apod. Konvertory v sobě mívají zabudované XML parsery.

Data zapsaná ve struktuře XML jsou vhodná pro výměnu a sdílení. XML je neutrální vůči platformám. Je to vlastně ASCII soubor s obecně známým kódováním, které je specifikováno v hlavičce souboru, je tedy možné kdekoliv ho správně přečíst a editovat v libovolném textovém editoru. Lze ho snadno integrovat do většiny aplikací (Musciano, 2000). Široké možnosti nabízí také odkazování v rámci XML. Je zajištěno jazyky XLink a XPointer. Dovolují odkazovat na vybranou část zdroje (např. určitou kapitolu či odstavec) (XPointer), vytvářet obousměrné odkazy pro vyjádření souvislosti dvou zdrojů, nebo udělat odkaz na několik zdrojů najednou (XLink) (Kosek, 1999).

3.4 RDF

Metadatové prvky nesou význam (sémantiku), která jednoznačně popisuje určitý objekt. XML slouží pro uložení těchto prvků. Především pro výměnu nebo sdílení metadatových záznamů pak slouží Rámec pro popis zdrojů (Resource Description Framework - RDF).

RDF je jazyk, jehož pomocí je možné reprezentovat informace o zdrojích v prostoru WWW. Nejde jen o zdroje, které je možné přes WWW přímo získat, ale i o zdroje, které je možné prostřednictvím WWW identifikovat. RDF je určen spíše pro strojové zpracování než pro zobrazování uživatelům. Jde o obecný rámec, takže

informace v RDF neztrácí význam ani při přenosu mezi různými aplikacemi (Manola, 2004).

Používá se jako obecný rámec pro kódování, výměnu a využívání metadat, které popisují objekty a reprezentují znalosti. RDF je vlastně další aplikací značkovacího jazyka XML (Miller, 1997). RDF umožňuje specifikovat sémantiku dat zapsaných v syntaxi XML ve standardizované podobě. RDF využívá jmenné prostory XML pro odkazy na definice formátů metadat. Určení metadatových formátů je nutné především pro interpretaci uložených dat.

RDF má široké využití. Lze ho užít pro popis obsahu a vztahů jedné i více webových stránek, pro popis jednoho zdroje i celé sbírky zdrojů, pro vyjádření a kontrolu vlastnických práv spojených se zdrojem. RDF usnadňuje vyhledávání, sdílení a výměnu informací. Nachází uplatnění nejen v knihovnictví a souvisejících oborech, ale také např. v obchodování pro popis jednotlivých výrobků, nebo pro vytváření záznamů o dodavatelích či zákaznících a jejich potřebách a preferencích do firemních databází (Manola, 2004).

RDF vznikl na půdě W3C (World Wide Web Consortium), které se o jeho rozvoj stará stále. Prvotními snahami W3C v této oblasti byl mechanismus PICS (Platform for Internet Content Selection), který byl tvořen jakousi stručnou charakteristikou zdroje v jeho hlavičce. Vznik PICS souvisel především se snahou umožnit filtrovat např. násilí, vulgarismy nebo sex na webových stránkách tak, aby se k nim nedostaly děti. PICS má velmi úzké a specifické použití. Jeho použití předpokládá, že všechny možné datové hodnoty jsou známy předem. Inspirací pro vznik RDF byl právě mechanismus PICS. Jeho rozšířením a úpravami vznikl RDF jako obecný rámec pro metadata nebo také obecný mechanismus reprezentace znalostí (Miller, 1997). Na vývoji RDF spolupracovalo mnoho organizací. Je ovlivněn zástupci komunity jak knihovnické, tak zástupci standardizační komunity a komunit pro reprezentaci znalostí a strukturování zdrojů. Do RDF byly vneseny prvky objektově orientovaného programování, stejně jako prvky modelování jazyka a tvorby databází.

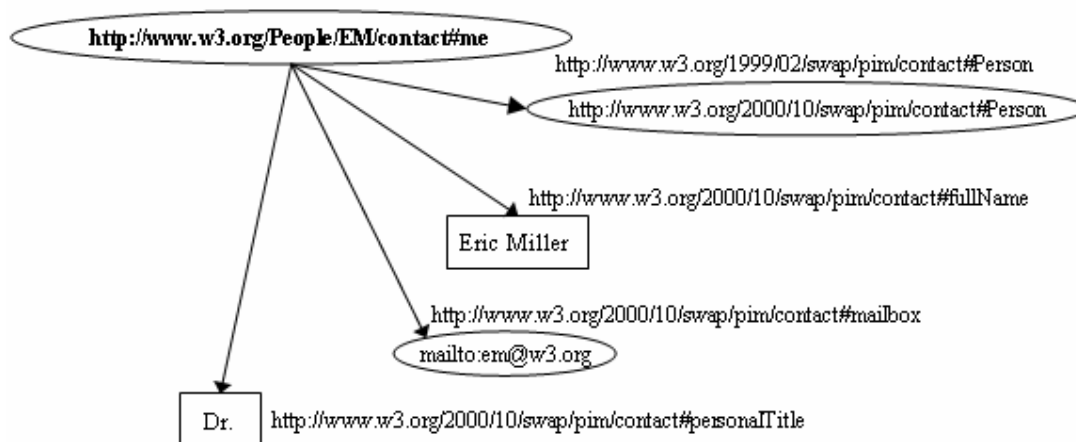
Mezi nejvýznamnější vlastnosti RDF patří vysoká expresivita RDF, portabilita, interoperabilita a dostupnost, podpora a otevřenost tohoto nástroje, kterou zaručuje konsorcium W3C (Vochozka, 2001). RDF lze užít pro popis všech zdrojů, které je možné identifikovat pomocí URI (Uniform Resource Identifier) a následně popsat zdroj pomocí jednotlivých vlastností a jejich hodnot (Manola, 2004). RDF

umožňuje kombinovat různá metadata a jejich slovníky. Při propojení více metadataových prvků zachovává jejich význam a vztahy mezi nimi.

RDF reprezentuje uzly (webové zdroje) identifikované URI. RDF vychází z toho, že objekt je možné popsat pomocí jednoduchých prohlášení. Každé prohlášení obsahuje tři části - určení objektu (subject), který je popisován, konkrétní vlastnosti (predicate), která se k objektu váže, a hodnoty (object), které vlastnost nabývá (Manola, 2004). Zdroj je tedy popsán pomocí souboru pojmenovaných vlastností a jejich hodnot. Hodnotou vlastnosti jednoho zdroje může být konkrétní řetězec znaků i odkaz na zdroj jiný, který má své vlastní vlastnosti a hodnoty (Miller, 1997). Časté jsou odkazy na určité hodnoty z řízených slovníků a také na prvky metadataových formátů. V takových případech jsou využívány jmenné prostory pro určení použitého slovníku či formátu.

Obrázek 3 tvoří soubor vlastností a jejich hodnot, které reprezentují osobu Erika Millera. Graf obsahuje několik URI, např. <http://www.w3.org/People/EM/contact#me> pro osobu Erika Millera, nebo <http://www.w3.org/2000/10/swap/pim/contact#Person> pro určení typu objektu (zde osoby). Identifikátory URI v RDF mohou být doplněny upřesňujícím volitelným identifikátorem části, který je uveden za znakem #. Takto doplněné URI se označuje URIref (URI reference). Jako ukázka vlastností a přiřazených hodnot slouží páry vlastností plné jméno (fullName) nebo titul (personalTitle) se svými hodnotami. Hodnotou může být také URIref. Potom se pro grafické znázornění hodnoty používá místo obdélníku ovál. Místo grafu je možné pro vyjádření vztahů mezi objektem, vlastností a hodnotou užít alternativní textový zápis, tzv. trojčlen (triple).

Obr. 3: RDF graf reprezentující osobu Erika Millera (Manola, 2004)



Aby bylo možné strojově zpracovávat a vyměňovat informace o objektech v RDF, byl na základě jazyka XML nadefinován jazyk RDF/XML. Ukázku záznamu v tomto jazyku obsahuje příklad 2, který odpovídá výše uvedenému grafu (Manola, 2004).

Příklad 2: RDF/XML záznam reprezentující osobu Erika Millera

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
  <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:fullName>Eric Miller</contact:fullName>
    <contact:mailbox rdf:resource="mailto:em@w3.org"/>
    <contact:personalTitle>Dr.</contact:personalTitle>
  </contact:Person>
</rdf:RDF>
```

Řadu informací je možné dále členit. I RDF umožňuje zaznamenat hierarchickou strukturu. Vlastností objektu může být adresa, kterou potřebujeme rozčlenit na město, ulici, poštovní směrovací číslo atd. Hodnotu adresy může zastupovat URIref, který se může stát sám objektem s několika vlastnostmi a hodnotami. Jeden zdroj se tak může stát hodnotou vlastnosti jiného objektu. Pokud není nutné odkazovat na jiný objekt pomocí URIref, přesto je potřeba rozčlenit hodnotu na více částí, vzniká tzv. prázdný uzel (blank node), ke kterému se jednotlivé části váží. V záznamu trojčlenů jsou pro tyto prázdné uzly používány identifikátory prázdných uzlů ve tvaru „_jméno“. Prázdné uzly mohou také zastupovat objekty, které nemají URI. Hierarchická struktura používaná v RDF zajišťuje, že vztahy v RDF budou pouze binární, dále dovoluje zpřesňovat prohlášení o objektech.

RDF nabízí řešení i pro popis skupiny objektů (např. studentů jednoho kurzu). Pro tyto případy nabízí několik předdefinovaných typů a vlastností. Jde především o slovník s třemi typy tzv. kontejnerů (container).

První kontejner lze přeložit jako vak (bag). Označuje se rdf:Bag. Užívá se pro skupinu objektů, které mohou být duplicitní a jejich pořadí ve skupině není určující.

Druhým kontejnerem je posloupnost (sequence) s označením rdf:Seq. Stejně jako u typu vak mohou být v případě posloupnosti objekty ve skupině duplicitní, ale tentokrát je pořadí objektů již důležité.

Kontejner alternativa (alternative) s označením rdf:Alt se používá pro skupinu různých hodnot jedné vlastnosti.

Při popisu je kontejner uvozen dle typu značkou `<rdf:Bag>`, `<rdf:Seq>` nebo `<rdf:Alt>` a jednotliví členové každého kontejneru značkou `rdf:_n`³. Aby nedošlo k duplicitě uvnitř kontejneru, do kódu se jednotlivé položky pro členy zapisují ve tvaru `rdf:li` a konkrétní čísla jsou vygenerována následně.

Jiným typem skupiny je sbírka RDF (RDF collection). Ta má, na rozdíl od kontejneru, prostředky, kterými určí omezenou skupinu objektů. Sběrka RDF má předdefinovaný typ `rdf:List` s hodnotami `rdf:first` a `rdf:rest` a předdefinovaným zdrojem `rdf:nil`. Každý ze členů skupiny je hodnotou vlastnosti `rdf:first` a tvoří jeden uzel. Seznam členů (uzlů) je propojen pomocí vlastnosti `rdf:rest`. Konec seznamu uzavírá vlastnost `rdf:rest` s hodnotou `rdf:nil` (Manola, 2004).

Pro případ, kdy je potřeba popsat pomocí RDF jiné prohlášení RDF a uvést např. údaje o datu vytvoření, autorovi apod., se používá slovník reifikace (reification). Výrok se v takovém případě stává sám popisovaným objektem a jsou o něm vytvářeny výroky (Manola, 2004; Pokorný, 2005). Pro popis se používá typ `rdf:Statement` s vlastnostmi `rdf:subject`, `rdf:predicate` a `rdf:object`.

Rozšířením RDF vzniká tzv. schéma RDF (zkráceně RDFS). RDFS poskytuje nástroje, jejichž pomocí lze vytvářet definice tříd a podtříd objektů, množin vlastností a jejich možných hodnot. RDFS dovoluje definovat hierarchickou strukturu. Jeden zdroj může náležet do více tříd. Třída představuje určité typové dělení objektů (`rdf:type`). Je tak možné nadefinovat různé třídy objektů (`rdfs:Class`) a jejich podkategorie (`rdfs:subClassOf`). Příkladem mohou být skupiny zvířata – savci – psi. Každá třída má nadefinovány vlastnosti. Podle schématu je následně možné RDF validovat. RDFS je samo o sobě zdrojem RDF a v záznamech je odkazováno pomocí `URIref` (Manola, 2004).

RDF má široké uplatnění. Používá se v aplikacích pro zápis a následnou manipulaci s informacemi o různých objektech. Vhodný je také pro metadatové formáty. Podrobněji se jazykem RDF a jeho použitím pro metadata zabývá diplomová práce Ondřeje Horsáka, která byla obhájena na Filozofické fakultě Univerzity Karlovy v roce 2005 (Horsák, 2005).

³ Písmeno „n“ představuje celou číslici vyšší než nula. Pro účely kontejneru se používají číslíce bez předřazené nuly, značka vypadá např. `rdf:_1`.

3.5 Protokol OAI-PMH



Základním cílem protokolu OAI-PMH je zajistit přístup k materiálům dostupným přes webový prohlížeč z různých archivů a repozitářů. Pomocí protokolu OAI-PMH je možné sklízet záznamy obsahující metadata z existujících archivů a pak je shromažďovat na jednom místě. Takto zpřístupněné záznamy mohou být v jakémkoliv metadatovém formátu. Je tedy možné sdružovat do jedné databáze záznamy v různých formátech.

OAI-PMH nenabízí nástroje pro vyhledávání v takto sklizených záznamech, ani nezajišťuje provázání nebo odkazy na popisované elektronické objekty. Dovoluje pouze přenést metadatové záznamy z různých míst na jedno centrální místo, kde mohou být tyto záznamy ukládány a sdružovány (Carpenter, 2003). Sklizení metadat probíhá jednoduchým způsobem. OAI-PMH definuje soubor dotazů a odpovědí, které si mohou zasílat dva servery. Dotazující server formuluje dotaz či příkaz a jako odpověď získá metadatový záznam, příp. více záznamů, uložený v XML (Summers, 2004).

Protokol OAI-PMH vytváří rámec pro stahování metadatových záznamů, který je nezávislý na typu aplikace. Používají ho dvě strany, které se účastní této komunikace, a to poskytovatel dat a poskytovatel služeb. Poskytovatel dat svoje data vystavuje a poskytovatel služeb tato data stahuje a poskytuje nad nimi přidané služby (Open Archives Initiative, 2004).

Protokol OAI-PMH je spojen s rozvojem digitálních archivů a knihoven. Každý digitální archiv volil vlastní metody ukládání zdrojů i jejich popisu, protokol, který dovolil autorům zasílat do archivu materiály a uživatelům vyhledávání a případné přidané služby. Tato různorodost technologií ale nepodporovala spolupráci mezi archivy, sdílení záznamů a rozšiřování služeb. Z tohoto důvodu se uskutečnilo v roce 1999 setkání představitelů digitálních archivů v Santa Fe v Novém Mexiku. Vyzvali širší odbornou komunitu ke spolupráci především v oblasti interoperability.

Klíčovou otázkou bylo, zda zvolit možnost vytvoření služby, která by prohledávala paralelně více digitálních archivů, nebo se přiklonit k možnosti stahování metadatových záznamů do jednoho nebo více centrálních archivů, kde by bylo zajištěno vyhledávání a další služby. S paralelním vyhledáváním měly digitální

archivy předchozí zkušenosti a tuto variantu zamítly jako nevhodnou pro požadovavné účely. Vznikl tak nový protokol pro sklizení metadatových záznamů OAI-PMH.

OAI-PMH měl především vyřešit, jaký protokol bude používán pro komunikaci, jaký bude základní metadatový formát, nároky na kvalitu minimálních záznamů a jak budou ošetřena a zajištěna autorská práva.

Nakonec byl pro komunikaci zvolen protokol HTTP pro zasílání žádostí a jazyk XML jako formát pro zasílání odpovědí. Základním metadatovým formátem se stal nekvalifikovaný Dublin Core, přípustný je však jakýkoliv jiný formát, který je možné zakódovat jazykem XML.

Verze 1 protokolu OAI-PMH byla testována a revidována. Verze 2 tohoto protokolu má širší záběr, slouží pro opakovanou výměnu metadatových záznamů. S verzí 1 není kompatibilní (Carpenter, 2003).

Protokol OAI-PMH je založen na protokolu HTTP, jehož prostřednictvím servery komunikují. Na jedné straně stojí poskytovatel služeb, který posílá příkazy na server poskytovatele dat, příp. servery více poskytovatelů dat.

OAI-PMH podporuje šest typů příkazů, které posílá server poskytovatele služeb. Může od archivu požadovat jeho identifikaci (příkaz *identify*), seznam dostupných metadatových formátů daného archivu (*ListMetadataFormats*), zjistit strukturu uložení záznamů v archivu (*ListSets*), stáhnout záznamy (*ListRecords*) nebo jejich hlavičky (*ListIdentifiers*) a získat vybraný záznam (*GetRecord*). Všechny tyto příkazy musí poskytovatel dat podporovat. Povolena je metoda get i post.

Opačným směrem, od poskytovatele dat k poskytovateli služeb, jsou zasílány základní informace o archivu, jeho struktuře, metadatových formátech, identifikátorech záznamů a samozřejmě samotné metadatové záznamy. Pokud jsou zasílány chybové hlášky, jsou založené na chybových hláškách HTTP.

Základními prvky, které jsou protokolem OAI-PMH definovány, jsou zdroj, jednotka, záznamy. Zdrojem je samotný popisovaný objekt fyzický nebo elektronický. Jednotka představuje virtuální obraz zdroje. Je to soubor všech záznamů o daném zdroji. Záznamy mohou být v různých formátech a jsou všechny identifikovány stejným identifikátorem zdroje. Formát identifikátoru musí odpovídat specifikaci URI (Uniform Resource Identifier).

Zasílané záznamy jsou textového charakteru a jsou uloženy v XML. Skládají se ze tří částí. Hlavička a samotná metadata jsou povinnou součástí zasílaného

záznamu. Třetí část „o“ (about) vypovídající o autorských právech a provenienci je volitelná.

Hlavička obsahuje jedinečný identifikátor jednotky, časové razítko poslední editace záznamu a případně specifikaci souboru, do kterého záznam v archivu patří. Vlastní metadatový záznam může být v libovolném metadatovém formátu. Základním formátem je nespecifikovaný Dublin Core. Povolený je však jakýkoliv jiný formát, který je možné zakódovat pomocí jazyka XML. Seznam dostupných formátů daného repozitáře je možné zjistit příkazem ListMetadataFormats. Prefix metadatového formátu ve jmenném prostoru XML určuje, v jakém metadatovém formátu záznam je (Carpenter, 2003).

Poskytovatelé dat mohou záznamy vystavovat samostatně nebo sdružené v souborech nebo sbírkách pro tzv. výběrové sklizení (selective harvesting). Pokud jsou záznamy sdružovány, musí být součástí odpovědi na dotazy ListIdentifiers, ListRecords a GetRecord také informace o existenci těchto souborů.

Výběrové sklizení může sklizeň metadat omezit jak pomocí příslušnosti do vybrané sbírky nebo více vybraných sbírek, tak pomocí časových známek záznamů. Časová známka se vztahuje k vytvoření, nebo editaci, příp. smazání záznamu o zdroji.

Na stránkách Iniciativy otevřených archivů je popis formátu OAI-PMH, kde je popsán každý příkaz a je zde také možné nalézt informace k implementaci tohoto protokolu (Open Archives Initiative, 2004).

3.6 Schéma METS

Schéma METS (Metadata Encoding and Transmission Standard, <http://www.loc.gov/standards/mets/>) nás opět přivádí přímo k metadatovým záznamům. Můžeme ho totiž zařadit mezi strukturální metadatová schémata.

Schéma METS bylo vyvinuto americkou Federací digitálních knihoven (Digital Library Federation). První verze METS byla zveřejněna v roce 2002. V současnosti je aktuální verzí schéma METS 1.5 z dubna 2005. Schéma je spravováno a dále rozvíjeno Kongresovou knihovnou. Jde o standard určený především pro použití v digitálních knihovnách pro kódování popisných, administrativních a strukturálních dat o digitálních objektech (Metadata Encoding and Transmission Standard, 1997).

Formát METS slouží především pro přenos a sdílení záznamů mezi různými systémy. Jeho cílem je sdružit v jednom záznamu všechna data, která se vztahují k jednomu informačnímu zdroji (Vojnar, 2005). METS představuje jakýsi kontejner, který spojuje dohromady všechny soubory, informace a záznamy, které se vztahují k jednomu zdroji. Souborů, které tvoří jeden zdroj, může být celá řada. Kromě souborů vztahujících se k samotnému zdroji nese záznam METS také informace technického charakteru (technická metadata), které jsou důležité pro následné použití souborů, např. formát souborů, informace o technologii použité při skenování, informace o případné kompresi apod. Obvykle jsou v záznamu METS vnořeny také záznamy s popisnými metadaty k danému objektu. Takových záznamů může záznam METS nést hned několik v různých popisných metadatových formátech (Coyle, 2005). Součástí záznamu mohou být administrativní metadata, která mohou např. ošetřovat přístup ke zdroji a stanovovat autorská práva spojená se zdrojem (Vojnar, 2005).

Dokument METS může sloužit jak pro popis objektů v repozitáři tak pro přenos záznamů o těchto objektech. Dokument se skládá ze sedmi základních částí, jimiž jsou hlavička METS, popisná metadata, administrativní metadata, seznam souborů objektu, strukturální mapa, strukturální odkazy a pravidla chování (Metadata Encoding and Transmission Standard, 1997). Každá část představuje vlastně jeden prvek, který je dále členěný do dceřinných prvků na dalších hierarchických úrovních.

Hlavička METS <metsHdr> obsahuje základní popisné údaje o samotném METS dokumentu. Jde především o datum vytvoření METS dokumentu, jeho aktualizaci, statutu, o jeho autorech a případně identifikátorech. Jak je vidět na příkladu 3, jednotlivé prvky mohou mít atributy k bližší specifikaci prvku. Obsah atributů bývá řízen slovníky.

Příklad 3: Ukázka hlavičky METS

```
<metsHdr CREATEDATE="2003-07-04T15:00:00" RECORDSTATUS="Complete">
  <agent ROLE="CREATOR" TYPE="INDIVIDUAL">
    <name>Jerome McDonough</name>
  </agent>
  <agent ROLE="ARCHIVIST" TYPE="INDIVIDUAL">
    <name>Ann Butler</name>
  </agent>
</metsHdr>
```


V části věnované popisným metadatovým záznamům mohou být přímo vnořeny metadatové záznamy, uveden odkaz na externí metadatový záznam, nebo použita kombinace předchozích možností. Pro každý metadatový záznam se opakuje prvek <dmdSec>. U každého odkazu na externí metadatový záznam je možné specifikovat typ identifikátoru (URN, URL, PURL, HANDLE, DOI, jiný), typ podle internetového standardu pro formát MIME (Multipurpose Internet Mail Extensions) a formát popisných metadat (MARC, MODS, EAD, VRA, DC, NISOIMG, LC-AV, TEIHDR, DDI, FGDC, jiný). Metadata přímo vnořená v záznamu musí být buď kódovaná v XML, nebo mohou být v libovolné binární či textové formě.

Sekce pro administrativní metadata <amdSec> zahrnuje v dokumentu METS několik základních typů informací o digitálním objektu. Jsou to technická metadata <techMD>, údaje o autorských právech <rightsMD>, zdrojová metadata o původních formách <sourceMD>, ze kterých byl digitální objekt odvozen, a metadata o digitálním původu objektů <digiprovMD>. Tagy pro typy informací jsou opakovatelné.

Následuje sekce <fileSec>, ve které jsou v prvku <fileGrp> uvedeny identifikátory jednotlivých souborů, které dohromady tvoří jeden objekt. Prvek je opakovatelný pro různé verze nebo formáty dokumentu.

Hierarchickou strukturu digitálního objektu zachycuje strukturální mapa <structMap>. Každé opakování prvku <div> v této části může kromě informací o popisované části obsahovat také ukazatel METS nebo ukazatel na soubor, které naznačují vztahy mezi jednotlivými soubory, objekty a metadaty v METS dokumentu. Strukturální odkazy <smLink> doplňují strukturální mapu o hypertextové odkazy mezi jednotlivými body této mapy. Odkazy jsou jednodušší variantou XML odkazů XLink.

Pravidla chování <behavior> obsahují doplňující informace o chování a případných programech, které souvisí s digitálním objektem, případně odkazují na externí pravidla chování a spouští různé programy.

Protože formát METS jako takový dává velkou volnost, mohou si organizace na jeho základě vytvářet podle určitých pravidel vlastní profily METS. Každý profil musí respektovat definovaná pravidla METS, ale může specifikovat také vlastní

pravidla a např. použití vybraných tezaurů a řízených slovníků pro určité prvky. Schválené profily METS eviduje Kongresová knihovna⁴.

Seznam projektů⁵, které plánují, nasazují nebo již používají formát METS, udržuje skupina digitálních knihoven SunSITE na Univerzitě v Berkeley, kterou podporuje také firma Sun (Metadata Encoding and Transmission Standard, 1997).

Obecně lze říci, že schéma METS je užitečný nástroj digitálních knihoven pro evidenci, výměnu a sdílení řady různých informací o digitálních objektech. Rychle si získává své místo mezi již existujícími standardy a technologiemi. Tím, že slouží mimo jiné pro seskupení popisných metadatových záznamů v různých formátech, napomáhá digitálním knihovnám a repozitářům v jejich spolupráci, což je dnes jedna z klíčových činností.

⁴ Seznam schválených profilů je dostupný na adrese <http://www.loc.gov/standards/mets/mets-registered-profiles.html>.

⁵ Seznam je dostupný na adrese <http://sunsite.berkeley.edu/mets/registry/>.

4 FORMÁT DUBLIN CORE A JEHO APLIKACE

4.1 Formát Dublin Core

Základy formátu Dublin Core (<http://www.dublincore.org/documents/dces/>), který je znám pod svým anglickým jménem a také pod označením „DC“, byly položeny na semináři, který v roce 1995 pořádala v Dublinu v Ohio americká knihovnická síť OCLC (Online Computer Library Center) spolu s Národním střediskem pro počítačové aplikace (National Centre for Supercomputing Applications, NCSA). Tento seminář měl pomoci řešit neuspokojivou situaci vyhledávání v síti Internet. Řešením pro kvalitnější a rychlejší vyhledávání relevantních informací v záplavě elektronických zdrojů publikovaných na Internetu měl být jednoduchý metadatový formát, který by používali především autoři a vydavatelé zdrojů (Chapesiuk, 1999). Na základě jednání na semináři vznikl soubor třinácti prvků, který dnes známe pod označením Soubor metadatových prvků Dublin Core (Dublin Core Metadata Element Set) (Bratková, 1999).

Nový metadatový formát neměl nahradit stávající systémy. Měl je doplnit a fungovat ve spolupráci s nimi. Východiskem byl předpoklad, že většina záznamů, které jsou generovány automaticky pro elektronické zdroje, neobsahuje dostačující údaje, aby mohl být popis použit. Naopak tvorba propracovaných záznamů typu MARC je velice finančně a časově náročná (Haigh, 1999). Formát DC měl představovat kompromis mezi těmito dvěma extrémů. Přestože je tento systém spojen především s popisem elektronických zdrojů, může sloužit i pro popis fyzických materiálů (DCMI, 1995c).

Metadatový formát DC je jednoduchý, sémanticky srozumitelný. Je flexibilní a překládá se do národních jazyků (Bearman, 1999). Délka žádného jeho pole není omezena. Popis lze aplikovat na jakýkoliv zdroj bez ohledu na nosné médium. Při popisu se uplatňuje tzv. pravidlo „1:1“. Jednotce by měl být přiřazen jeden záznam. Nicméně stanovit jednotku není vždy zcela jednoduché. Za jednotku může být považována jedna sbírka, webová stránka nebo jejich komplex, časopis, jednotlivý článek z časopisu, případně i menší část informačního zdroje.

Formát Dublin Core se stále vyvíjí a reaguje na nové změny a požadavky. Vývoj můžeme sledovat na www stránkách určených tomuto formátu, v rámci

elektronických diskusí vývojových skupin a na seminářích a konferencích, které pořádá Iniciativa pro metadatový formát Dublin Core (Dublin Core Metadata Initiative, DCMI). DCMI je virtuální komunitou odborníků pro metadata. Je podporována knihovnickou sítí OCLC (Online Computer Library Center), Britskou kancelář pro knihovní a informační sítě (UK Office of Library and Information Networking, UKOLN), americkou Národní vědeckou nadací (National Science Foundation, NSF) a dalšími významnými institucemi. DCMI je zodpovědná za vývoj, standardizaci, fungování a provoz souboru prvků Dublin Core (Bearman, 1999).

V rámci DCMI funguje řada pracovních skupin (Working Groups), které řeší specifické problémy (DCMI, 1995g). Pracovní skupiny vznikají a zanikají podle aktuálních praktických potřeb a problémů, které je třeba vyřešit. V současné době existují např. skupiny pro řešení architektury, tvorbu uživatelských příruček a manuálů nebo skupiny pro implementaci popisu Dublin Core do vybraných oblastí jako jsou vzdělávání, životní prostředí nebo vládní sféra. V průběhu existence DCMI vznikala a byla uzavírána řada skupin, které již svoji funkci splnily. Lze připomenout zaniklé pracovní skupiny, které se zabývaly např. otázkami věcného popisu, datových modelů, pravidla 1:1 apod. (DCMI, 1995d).

Iniciativa Dublin Core pravidelně pořádá konference. Projednává se na nich např. směr rozvoje formátu DC, specifikace syntaxe pro přenos dat, způsob vytváření průvodců a manuálů pro uživatele, identifikace mechanismů pro rozšíření užití formátu a specifikace rámce pro sladění formátu DC s dalšími formáty metadat (Haigh, 1999). Opakovaně byla na seminářích potvrzována nezbytnost některých požadavků na formát, jako jsou např. opakovatelnost a volitelnost prvků, nezávislost na pořadí prvků, robustnost, formálnost, úspornost, jednoduchost, jednoznačnost a srozumitelnost, možnost zúžit význam prvku pomocí kvalifikátorů, požadavek rozšíření formátu pro mezinárodní užití. Také byl stanoven princip, jakým způsobem uvádět metadata v rámci HTML zdroje (Burnard, 1996).

Důležitou otázkou byla tvorba základních pravidel architektury metadat. Dnes je označována jako Warwický rámec (Warwick Framework). Warwický rámec byl konceptuálním modelem pro metadata. Rozlišoval několik typů metadat (Haigh, 1999):

- *jednoduchý soubor metadat* - záznam jednoho zdroje, soubor metadat určený k přesnému účelu. Jeden balíček metadat může obsahovat metadata DC, další balíček bude záznam MARC atp. Každý balíček může být označen jako objekt, ale objektem může být i celý soubor metadat, nebo jen jeden prvek, nebo odkaz na nějaký externí objekt.
- *nepřímý soubor metadat* - odkaz na zdroj, který obsahuje metadata
- *sbírka více balíčků metadat (tzv. kontejner)* - hromaděním jednotlivých balíčků vzniká architektura nazvaná kontejner.

Na třetím semináři v září 1996 v americkém Dublinu byly k základní skupině původně třinácti prvků přidány další dva. Po změně uspořádání prvků a dílčích aktualizacích byl v roce 1998 tento soubor metadat přijat pod označením RFC 2413 a sloužil jako internetová norma (Bratková, 1999).

Čtvrtý seminář, který se konal v roce 1997 v Austrálii, byl provázen rozporem mezi tzv. minimalisty a strukturalisty. Tyto dvě skupiny vznikly jako protipól možnosti rozšíření formátu DC. Minimalisté hájí zachování co největší jednoduchosti formátu. Tvrdí, že jednoduchost a sémantická jednoznačnost jsou důležité proto, aby mohli záznamy metadat připravovat nezaškolení tvůrci webovských a vůbec elektronických zdrojů, a také proto, aby mohla být zajištěna interoperabilita mezi různými systémy. Naopak strukturalisté se snaží vyzdvihnout potřebu rozšíření prvků nebo alespoň jejich specifikace. Poukazují na nutnost rozšíření prvků a možnost zpřesnit popis pomocí kvalifikátorů. To by podle nich umožnilo lepší výsledky vyhledávacích strojů. Strukturalisté navrhli tři typy kvalifikátorů:

- kvalifikátor typu „TYPE“ (typ) - zužuje význam daného prvku
- kvalifikátor typu „SCHEME“ (schéma) - vztahuje prvky k nějakému obecně známému schématu, podle kterého má být vysvětlován jejich význam
- kvalifikátor typu „LANGUAGE“ (jazyk) - určuje, v jakém jazyce je zapsána hodnota prvku

(Bartošek, 1999)

Následující semináře přinesly diskuse mimo jiné o vztahu mezi Dublin Core a protokolem Z39.50, úvahy o popisu komplexu zdrojů se smíšenými médii (Haigh, 1999). Dalšími problémy byla formalizace procesů v souvislosti s odlišnými požadavky různých oborů, otázky nahrazení neformální konvence kódování v HTML formální specifikací, nutnost přesně vymezit mechanismy kvalifikátorů a vytvořit kontrolované slovníky autorit. Vedly se úvahy o využití metadat ve vládní sféře a státní správě. V roce 2000 byla pozornost věnována také jazykovým mutacím formátu a možnostem jejich převodu (DCMI, 1995b).

V roce 2003 se konala setkání v Seatlu. V Seatlu byl hlavním tématem tzv. abstraktní model metadat Dublin Core (Dublin Core Abstract Model), který reprezentuje metadatový popis Dublin Core. Na jeho základě je pak možné porovnávat metadata Dublin Core s jinými metadatovými schématy.

Hlavními tématy konference v Číně v Šanghaji v roce 2004 byly otázky metadatových aplikací a především jejich interoperabilita, která zahrnuje řešení jazykových bariér, kulturních bariér i odlišností zpracovatelské praxe. Příspěvky se soustředily také na budoucnost metadat, uživatele metadat a způsoby uchovávání a správy metadat a elektronických zdrojů.

V roce 2005 hostoval setkání příznivců metadat Dublin Core španělský Madrid. Diskuse byly zaměřeny především na vytváření a udržování řízených slovníků v souvislosti s popisem webovských zdrojů a využití různých třídění a taxonomií (DCMI, 2005).

V říjnu 2006 proběhla konference v Manzanillu v Mexiku. Mezi klíčová témata patřila interoperabilita, řízené slovníky a využití metadat pro výuku a vzdělávání nejen na univerzitách a ve školách, ale také v paměťových a vzdělávacích institucích typu knihoven nebo muzeí (DCMI, 1995a).

Jak již bylo zmíněno, patnáct metadatových prvků bylo v roce 1998 uznáno internetovou normou IETF RFC 2413. Dalším úspěchem na poli standardizace se stala informační norma IETF RFC 2731 nazvaná Kódování metadat DC v HTML, která slouží pro definici způsobu, jakým se metadata vkládají do jazyka HTML. V září 2001 ratifikoval Americký národní úřad pro normalizaci (American National Standards Institute, ANSI) verzi 1.1 jako národní americký standard ANSI Z39.85. Standard připravila americká nezisková asociace NISO (National Information Standard Organization). V roce 2003 byl soubor Dublin Core schválen Mezinárodní

organizací pro standardizaci ISO (International Standard Organization) jako ISO 15836:2003. Stal se tak mezinárodní normou (ISO 15839:2003).

4.1.1 Popisné jednotky

Iniciativa Dublin Core ponechává na rozhodnutí každé instituce, co bude považovat za popisnou jednotku. Obecně ale doporučuje použít princip 1:1 a použít vždy jeden záznam pro jednu manifestaci díla (Hillmann, 2004).

4.1.2 Metadatové záznamy

Formát Dublin Core obsahuje patnáct základních prvků. Jde o:

- *Datum* - časové určení vytvoření, zpřístupnění nebo jiných důležitých změn textu
- *Formát* - typ média, rozměry zdroje, určení hardware či software apod.
- *Identifikátor zdroje* - jednoznačně vymezuje a určuje zdroj, např. URL, DOI nebo ISBN
- *Jazyk* – jazyk zdroje vyjádřený dvouznačkovým kódem
- *Název* - jméno dané zdroji, jméno, pod nímž je zdroj oficiálně znám
- *Pokrytí* - rozsah nebo záběr obsahu zdroje, může to být časové či prostorové umístění, jurisdikce apod., k vyplnění tohoto pole slouží řízený slovník
- *Popis* - obsah zdroje vyjádřený pomocí abstraktu, obsahu, volného textu nebo odkazu na jiné vyjádření obsahu
- *Předmět a klíčová slova* - obsah zdroje, doporučuje se vybírat hodnoty prvku z řízených slovníků nebo schémat třídění
- *Příspěvatel* - sekundární odpovědnost za obsah zdroje
- *Správa autorských práv* - duševní vlastnictví, autorské právo a jiná vlastnická práva
- *Tvůrce* - osoba nebo organizace primárně odpovědná za vytvoření obsahu zdroje
- *Typ* - povaha nebo druh obsahu zdroje, popisuje obecné kategorie, funkce nebo druhy obsahu, opět s výběrem hodnoty z řízeného slovníku
- *Vydavatel* - osoba či organizace zodpovědná za zpřístupnění obsahu
- *Vztah* - odkazuje na příbuzný zdroj
- *Zdroj* - odkazuje na původní zdroj, ze kterého je popisovaný objekt odvozen (DCMI, 2006b; Masarykova univerzita, 2004)

Tab. 1: Základní prvky s kvalifikátory a schémata zápisu

Prvek	Kvalifikátor	Schéma zápisu
Datum (date)	Dostupný (available) Vytvořen (created) Datum schválení (dateAccepted) Datum copyrightu (dateCopyrighted) Datum odevzdání (dateSubmitted) Vydán (issued) Modifikovaný (modified) Platný (valid)	DCMI období W3C-DTF
Formát (format)	Rozsah (extent) Médium (medium)	IMT
Identifikátor zdroje (identifier)	Bibliografická citace (bibliographicCitation)	URI
Jazyk (language)		ISO 639-2 RFC 1766 RFC 3066
Název (title)	Alternativní (alternative)	
Pokrytí (coverage)	Prostorový (spatial) Časový (temporal)	DCMI Box ISO 3166 DCMI bod TGN
Popis (description)	Abstrakt (abstract) Obsah (tableOfContents)	
Předmět a klíčová slova (subject)		DDC, LCC, LCSH, MeSH NLM, UDC
Prispěvatel (contributor)		
Správa autorských práv (rights)	Přístupová práva (AccessRights) Licence (license)	
Tvůrce (creator)		
Typ (type)		Slovník typů DCMI
Vydavatel (publisher)		
Vztah (relation)	Odpovídá čemu (conformsTo) Má formát (hasFormat) Obsahuje část (hasPart) Má verzi (hasVersion) Je formátem (isFormatOf) Je částí (isPartOf) Je odkazován (isReferencedBy) Je nahrazen (isReplacedBy) Je požadován (isRequiredBy) Je verzí (isVersionOf) Odkazuje (references) Nahrazuje (replaces) Vyžaduje (requires)	URI
Zdroj (source)		URI

Každý prvek je možné specifikovat pomocí kvalifikátoru. Ve formátu Dublin Core jsou kvalifikátory dvojího typu - kvalifikátor zpřesňující význam celého prvku a schéma zápisu (obvykle všeobecně známé schéma).

První typ kvalifikátoru zužuje význam či blíže specifikuje prvek, se kterým je použit. Prvek neztrácí svůj obecný význam, ale význam se pouze zužuje. Pokud systém neumí tento kvalifikátor interpretovat, lze ho ignorovat a převzít obecný smysl prvku.

Schéma zápisu poukazuje na specifikaci schématu, v jehož rámci se má interpretovat prvek a jeho hodnota. Tímto schémem může být řízený slovník, formální notace či jiná pravidla. Pokud vyhledávací stroj nerozumí zápisu obsahu, obsah je srozumitelný lidskému čtenáři (jde např. o standardní formu zápisu data 2006-02-13). Odkaz může být učiněn pouze na existující schéma (DCMI, 2000).

Podle toho, zda jsou nebo nejsou použity kvalifikátory, rozlišujeme popis kvalifikovaný a nekvalifikovaný. Kvalifikátory mohou v některých případech zapříčinit nekonzistenci a zhoršenou výměnu informací. Jsou však vhodné pro upřesnění popisu. V každém případě je možné kvalifikátory ignorovat, čímž bude sice omezena přesnost a specifika popisu, ale nedojde k celkové nesrozumitelnosti záznamu. Základní význam popisu zůstane zachován (DCMI, 1995c). Tabulka 3 shrnuje všechny základní prvky Dublin Core a uvádí kvalifikátory, které je možné s těmito prvky použít (DCMI, 2006a; Masarykova univerzita, 2005).

Kromě základního souboru patnácti prvků Dublin Core je možné použít několik dalších prvků. Jsou uvedeny v tabulce 2 (DCMI, 2006a; Masarykova univerzita, 2005). Byly povoleny v záznamech DC s určením, zda jsou doporučené nebo vyhovující pro použití v popisu. Obvykle prvek navrhl pracovní skupina DC nebo určitá komunita, která potřebovala evidovat některé údaje nad rámec základních prvků DC.

Tab. 2: Další povolené prvky formátu Dublin Core

Prvek	Kvalifikátor
Metoda přírůstku (accrualMethod)	
Periodicita přírůstku (accrualPeriodicity)	
Strategie výběru přírůstků (accrualPolicy)	
Uživatelské určení (audience)	Úroveň vzdělání (educationLevel) Zprostředkovatel (mediator)
Vzdělávací metoda (instructionalMethod)	
Původ (provenance)	
Držitel práv (rightsHolder)	

V rámci popisu pomocí metadat Dublin Core je možné použít řadu rejstříků a tezaurů. Většina z nich je obecně uznávaná, spravovaná a používána jinými organizacemi. Některé řízené slovníky připravila iniciativa Dublin Core přímo pro použití v metadatovém formátu DC. Příkladem může být řízení slovník pro určení typu zdroje Slovník typů DCMI:

- Sběrka (collection)
- Soubor dat (dataset)
- Událost (event)
- Obraz (image)
- Interaktivní zdroj (interactiveResource)
- Pohyblivý obraz (movingImage)
- Fyzický objekt (physicalObject)
- Služba (service)
- Software (software)
- Zvuk (sound)
- Statický obraz (stillImage)
- Text (text)

(DCMI, 2006a; Masarykova univerzita, 2005)

Metadata Dublin Core se dle doporučení mají spíše uchovávat v oddělených .rdf souborech než přímo vkládat do HTML, příp. XML stránky (DCMI, 1995c). Jako praktická ukázka záznamu ve formátu Dublin Core může posloužit příklad 4. Obsahuje jednoduchý záznam z webových stránek DCMI, na kterém jsou vidět dvojice vlastností a hodnot vztahujících se k jednomu prvku. Úvodní řádka odkazuje na specifikaci formátu.

Příklad 4: Záznam ve formátu Dublin Core

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.title" content="DCMI Documents" />
<meta name="DC.description" content="Lists of DCMI recommendations, proposed recommendations, superseded recommendations, process documents, recommended resources and important working drafts." />
<meta name="DC.date" content="2005-06-13" />
<meta name="DC.format" content="text/html" />
<meta name="DC.contributor" content="Dekkers, Makx" />
<meta name="DC.contributor" content="Baker, Thomas" />
<meta name="DC.language" content="en" />
<meta name="DC.publisher" content="Dublin Core Metadata Initiative" />
```

Kvalifikovaný záznam Dublin Core zapsaný v XML je v příloze 1. Záznam je obsáhlejší, má složitější strukturu než uvedený jednoduchý záznam v příkladu 4 a jsou v něm použity kvalifikátory i odkazy na některá schémata.

Soubor prvků Dublin Core zachovává tři základní principy. První princip „dumb-down“ vyjadřuje možnost použít kterýkoliv ze základních prvků spolu s kvalifikátorem nebo bez něho. Pokud systém nepodporuje použití kvalifikátorů, stejně bude obsah záznamu srozumitelný. Druhé pravidlo „1:1“ doporučuje použít vždy jeden záznam pro jednu manifestaci díla. Třetí princip spočívá ve vhodném výběru hodnot pro jednotlivé prvky, kdy by tvůrce metadat měl mít na zřeteli především cílového uživatele a jeho možné uvažování při vyhledávání (Hillmann, 2004).

Metadatový formát Dublin Core nemá nahradit ostatní metadatové formáty, naopak funguje v součinnosti s nimi. Organizace mohou formát modifikovat podle stanovených pravidel tak, aby splňoval jejich potřeby. Je možné přidávat vlastní prvky nebo používat prvky z jiných formátů. V případě, kdy je některý prvek převzat z jiného schématu metadat, má svůj vlastní prefix a není uvozen prefixem „DC.“ (Haigh, 1999).

Iniciativa Dublin Core na svých stránkách aktualizuje doporučení pro vytváření metadatových záznamů a zveřejňuje další manuály a podklady spolu s odkazy na existující pomůcky pro vytváření metadatových záznamů.

4.1.3 Shrnutí informací k formátu Dublin Core

Metadatový formát Dublin Core si již bezpochyby upevnil své postavení ve světě popisu elektronických (a nejen elektronických) informačních zdrojů. Iniciativa DCMI pořádá řadu odborných akcí, spolupracuje s řadou institucí a vyzývá i další organizace ke spolupráci. Rozvíjí se soubor metadatových prvků. Přibývají další prvky, které se začleňují do kvalifikovaného popisu a formát se dále vyvíjí. V současnosti jsou prvky formátu Dublin Core přeloženy zhruba do pětadvaceti jazyků, mezi jazykovými mutacemi nechybí ani česká mutace a existují také mutace japonská a čínská. Patrná je snaha o harmonizaci popisu nejen mezi jednotlivými institucemi a zeměmi, ale také například v určitých tematických oblastech (DCMI, 1995e).

I nadále lze očekávat, že v rámci iniciativy DCMI budou fungovat pracovní skupiny, které budou reagovat na nové možnosti ve světě informačních zdrojů, budou zvažovat možné přínosy nových technologií a pravidel pro popis a následné vyhledávání, budou se zabývat spoluprácí a kooperací informačních institucí.

Metadatový formát Dublin Core se stejně jako jiné další formáty stal vzorem pro řadu dalších formátů a nástrojem pro popis zdrojů v různých projektech a systémech. Některé převzaly formát jako takový, jiné ho použily pro vytvoření vlastního formátu, který z formátu DC vychází. Výjimkou nejsou ani projekty, které používají nebo alespoň nabízí export záznamu v různých formátech. Jedním z nich je obvykle formát DC. Používá se také v řadě projektů jako standardizovaný formát pro výměnu a sdílení záznamů v různých formátech, např. v OAI-PMH.

Některé projekty přinesly řadu nástrojů pro snadnější vytváření metadatových záznamů. Například v rámci severského Severského metadatového projektu (Nordic Metadata Project), který probíhal ve dvou návazných projektech v letech 1996-2000 pod záštitou informační sítě NORDINFO, byl připraven pro skandinávské země interaktivní formulář pro tvorbu metadat ve formátu DC za účelem sdílení záznamů metadat. Také byl vytvořen generátor URN (Uniform Resource Names - unifikovaná jména zdrojů) a experimentální konvertor pro převod metadat do formátů typu MARC (Nordic Metadata Projects, 2000).

V České republice se také zvýšila potřeba popisu a vyhledávání elektronických zdrojů a pozornost se zaměřila na formát Dublin Core. První aktivity vyvinulo Knihovnicko-informační centrum Masarykovy univerzity v Brně, které v roce 2000 zahájilo spolupráci s iniciativou Dublin Core a přeložilo prvky formátu DC do češtiny. V roce 2006 byl překlad aktualizován (Masarykova univerzita, 2004).

Na příkladech několika vybraných projektů a systémů (viz. podkapitoly 4.2 - 4.6), které používají formát Dublin Core si demonstrujeme, jak rozmanitě může být formát použit k různým účelům. Každá komunita, které formát přijala, přistoupila k jeho používání z hlediska svých potřeb. Některé komunity ponechaly formát v původní nezměněné formě, jiné si formát uzpůsobily pro své účely, přidaly některé prvky, nebo alespoň využívají vlastní kódovací schémata.

4.2 Formát AGLS



Na základě studií, které se zaměřovaly na výsledky vyhledávání vyhledávacích strojů na Internetu ve srovnání s použitím metadat, rozhodla Australská vláda o vytvoření metadatového formátu pro popis vládních zdrojů. Metadatový formát dostal název Ukazatelová služba australské vlády, zkráceně AGLS (Australian Government Locator Service, http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html). Formát AGLS vznikl na základě metadatového formátu Dublin Core na přelomu let 1997 a 1998 (National Archives of Australia, 2000). V prosinci 2002 byl publikován jako australská norma AS 5044.

Původně měl AGLS vycházet z metadatového formátu GILS⁶, který je využíván pro popis zdrojů americké vlády, a měl nést název AUSGILS. Ukázalo se však, že formát GILS má příliš složitou strukturu, což nejen působí problémy, ale také vyžaduje značné časové a finanční náklady na tvorbu záznamů. Proto byl na základě konzultací s odborníky později AGLS vyvíjen na bázi formátu Dublin Core. Je to formát velice jednoduchý a umožňuje tvorbu záznamu přímo v době vzniku zdroje, ať jde o popis tvořený autorem zdroje nebo automaticky tvořený popis. Dalším důvodem výběru formátu Dublin Core byl výhled, že tento formát nezanikne, nýbrž se bude dále rozvíjet jako mezinárodně uznávaný formát (Cunningham, 1998).

Podobně jako Dublin Core je i AGLS schopen dalšího rozvoje a umožňuje v případě potřeby konkrétní organizace začlenit do popisu další prvky, které však musí mít stejnou sémantiku jako prvky AGLS. Podobně je možné zacházet i s kvalifikátory. Stejně jako u formátu Dublin Core musí mít prvky jasný význam, i pokud nejsou užity kvalifikátory (National Archives of Australia, 2000).

Ačkoliv by se mohlo zdát, že jsou oba formáty velmi podobné, je mezi nimi možné najít řadu rozdílů. AGLS byl oproti formátu Dublin Core rozšířen o několik prvků. Další rozdíl můžeme spatřovat v povinných prvcích a dále ve způsobu použití kvalifikátorů. Pomocí kvalifikátorů ve formátu AGLS může např. tvůrce vyjádřit,

⁶ Formátem GILS se zabývá diplomová práce Radomíry Cvrčkové (Cvrčková, 2000).

že v prvku použil výraz z tezauru nebo jiného řízeného slovníku, nebo upřesnit formát, který použil pro zápis data. Formát Dublin Core může popisovat libovolný objekt, ale primárně je určen k popisu elektronických informačních zdrojů. Formát AGLS se používá nejen k popisu online zdrojů, ale ve velké míře také k popisu zdrojů offline, které zahrnují i muzejní objekty, obrazy, tištěné zdroje. Je také uzpůsoben tak, aby kromě popisu objektů dokázal popsát také služby a organizace.

Australská vláda doporučuje všem podřízeným organizacím popisovat elektronické zdroje na vládní úrovni federální, státní i lokální. Cílem je zlepšení dostupnosti a interoperability vládních zdrojů a služeb právě pomocí standardizovaného popisu založeného na metadatech. Uživatel by měl pomocí vyhledávacích strojů Internetu snadněji najít informace, které hledá. Vývoj formátu AGLS odráží pojetí Internetu australskou vládou především jako prostoru pro komunikaci s občany (Cunningham, 1998). Běžné vyhledávací stroje Internetu zatím vyhledávání dle metadat nepodporují tak, jak se očekává. Předpokládá se, že do budoucna bude možné zadat, aby agent vyhledával ve zvoleném metadatovém prvku stanovenou hodnotu. Kromě vyhledávacích agentů je také možné použít k vyhledávání webové rozhraní (<http://www.australia.gov.au/>), které podporuje vyhledávání podle metadat AGLS (National Archives of Australia, 2000).

Všechny vládní úřady a další organizace nepoužívají AGLS stejným způsobem. Musí si vyvinout některé postupy a pravidla dle svých specifických potřeb. Mezi některá rozhodnutí každé organizace patří volba metadatových prvků, výběr zdrojů k popisu, uložení metadat, užití syntaxe (RDF, XML,...), výběr používaného tezauru apod. (Cunningham, 1998). Pokud některá organizace provádí úpravy formátu AGLS tak, aby lépe splňoval její potřeby, musí zachovávat několik základních pravidel. Především všechny stávající prvky AGLS musí mít stejný význam, povinné prvky musí také zůstat zachovány a význam všech přidávaných kvalifikátorů musí být v souladu s již existujícími prvky a kvalifikátory (National Archives of Australia, 2000b).

Jako orgán, který spravuje formát AGLS, byla vybrána instituce Národní australské archivy (National Archives of Australia, NAA). Mezi hlavní úkoly Národních australských archivů patří vedení pracovních skupin AGLS a koordinace vývoje formátu, správa uživatelských manuálů, pomoc při zavádění standardu do vládních agentur a úřadů, udržování www stránek. Sledují také vývoj formátů, které navazují na formát AGLS. Národní archivy kontrolují také, zda ostatní

organizace používají formát AGLS správně. Je-li tomu tak, vystavují tyto organizace na svých webových stránkách logo AGLS. Kromě Národních australských archivů se podílejí na vývoji AGLS i další organizace. Mezi nejvýznamnější patří Národní kancelář pro informační ekonomii (National Office for the Information Economy, NOIE) a Úřední rada online (Online Council Officials) (National Archives of Australia, 2000).

Ostatní australské metadatové formáty jsou založené buď na formátu Dublin Core nebo na formátu AGLS. Jednou z výjimek je metadatový formát ANZLIC (Územní informační rada Austrálie a Nového Zélandu, Australia and New Zealand Land Information Council) pro popis geografických datových souborů. Pracovníci AGLS připravili konverzní tabulku pro převod záznamů ANZLIC do formátu AGLS.

Mezi nejvýznamnější formáty, které vycházejí z formátu AGLS, patří Business Entry Point (BEP) nebo HealthInsite. Oba jsou australské. První slouží pro popis obchodních a politických zdrojů a druhý, jak napovídá jeho název, pro popis zdrojů z oblasti medicíny a zdraví (National Archives of Australia, 2000).

4.2.1 Popisné jednotky

Záleží na rozhodnutí každé instituce, jaký stupeň popisu pro své zdroje zvolí. Doporučení AGLS umožňují tvorbu záznamu na úrovni jednotky, sbírky i celého souboru zdrojů. Každá instituce by měla zvážit stupeň popisu především podle očekávaných uživatelů a jejich předpokládaných potřeb. Podobným způsobem vybírá každá organizace také zdroje, které bude pomocí metadat popisovat. Není nutné popisovat každý jednotlivý zdroj, ale spíše vybrané zdroje (National Archives of Australia, 2000). Nepopisují se pouze zdroje elektronické, ale AGLS slouží také k popisu fyzických dokumentů (Wilson, 2002c).

Jako doplněk k používání formátu AGLS pro popis objektů doporučuje Národní australský archiv používání také tzv. A-Core metadat, které představují administrativní metadata o metadatach a jejich významem je nést informace o vytvoření metadatového záznamu objektu. Jsou to údaje o tom, kdy byl metadatový záznam vytvořen a kdo byl jeho tvůrce (National Archives of Australia. 2002b).

4.2.2 Metadatové záznamy

AGLS je soubor devatenácti popisných prvků. Převzal prvky metadatového formátu Dublin Core a přidal další dle vlastních potřeb. Prvky přidané do formátu AGLS jsou *dostupnost*, *funkce*, *mandát* a *uživatelské určení* (National Archives of Australia, 2000).

Prvek *dostupnost* byl přidán především kvůli funkci lokační. Jeho primární funkcí není pouze nalezení zdroje, ale zároveň i zajištění informací nutných k dodání vybraného zdroje. Obsahuje informace, jak zdroj přímo získat, nebo alespoň kontakty, na kterých lze požádat o jeho zaslání.

Prvek *funkce* je považován za důležitý především z hlediska zveřejnění vládních zdrojů. V rámci tohoto prvku jsou specifikovány funkce a úkoly, ke kterým může být zveřejněný zdroj přiřazen. Jde o jedno z hledisek využívaných při vyhledávání. Pomocí tohoto prvku se také zdroje zařazují do celé sítě vládních zdrojů (Cunningham, 1998).

Prvek „uživatelské určení“ vymezuje okruh původně zamýšlených uživatelů zdroje.

Prvek *mandát* nese informace o rozhodnutí, na jehož základě byl popisovaný zdroj vytvořen. Jde o různá právní nařízení, předpisy, zákony a výnosy (National Archives of Australia, 2000).

AGLS je formát jednoduchý, flexibilní a dynamický. Část prvků je povinná, ostatní jsou volitelné. Povinné prvky jsou autor, vydavatel, název, datum, předmět nebo funkce, identifikátor nebo dostupnost. Přítomnost ostatních prvků v záznamu závisí na rozhodnutí každé organizace. Pořadí prvků není v záznamech AGLS rozhodující.

Podobně jako Dublin Core, i AGLS definuje pro každý prvek několik vlastností a pravidel používání. Základní je název prvku a zda je povinný či volitelný, je uvedena jeho stručná definice, soubor termínů, které mohou být v rámci prvku použity. Dále je prvek možné omezit pomocí řízených slovníků a schémat zápisu, případně existují hodnoty, kterými lze rozšířit strukturu prvku, nejsou-li užita externí schémata a řízené slovníky. Stejně jako u formátu Dublin Core umožňuje i AGLS nekvalifikovaný popis. Prvek má svoji přednastavenou hodnotu.

Každý prvek je opakovatelný, není omezena délka hodnoty prvku ani použité typy znaků. Nezáleží na jejich pořadí v záznamu. AGLS používá standardně

angličtinu, nicméně není stanoveno žádné omezení týkající se jazyka (National Archives of Australia, 2000). Prvky formátu AGLS s kvalifikátory jsou uvedeny v tabulce 3 (National Archives of Australia, 2002a).

V rámci AGLS platí také doporučení, jak zapisovat vstupní hodnoty. Např. u prvku tvůrce platí doporučení zapisovat u jmen osob první příjmení a za čárkou jméno osoby. Pro datum vytvoření je zase doporučováno použít formu podle ISO 8601 a zakódovat datum do podoby YYYY-MM-DD (National Archives of Australia, 2002a).

Tab. 3: Prvky formátu AGLS s kvalifikátory

Prvek	Kvalifikátor	Význam kvalifikátoru
Tvůrce (creator)		
Datum (date)	Vytvoření (created)	Datum vytvoření zdroje
	Modifikace (modified)	Datum modifikace zdroje
	Platnosti (valid)	Rozmezí dat pro platnost zdroje
	Zveřejnění (issued)	Datum oficiálního zveřejnění zdroje v jeho současné podobě
Název (title)	Variantní (alternative)	Alternativní název zdroje – např. zkratka, akronym
Dostupnost (availability)		
Funkce (function)		
Identifikátor (identifier)		
Vydavatel (publisher)		
Předmět (subject)		
Příspěvatel (contributor)		
Popis (description)		
Formát (format)	Rozsah (extent)	Fyzické rozměry, velikost souboru, doba trvání skladby apod.
	Nosič (medium)	Materiál nebo fyzický nosič zdroje
Jazyk (language)		
Práva (rights)		
Zdroj (source)		
Typ zdroje (type)	Kategorie (category)	Nabývá hodnot služba, dokument nebo agentura
	Stupeň agregace (aggregationLevel)	Stupeň sdružení jednotek. Nabývá hodnot jednotka nebo sbírka
	Typ dokumentu (documentType)	Forma zdroje
	Typ služby (serviceType)	Typ služby

Prvek	Kvalifikátor	Význam kvalifikátoru
Vztah (relation)	Je verzí (isVersionOf)	Odkaz na zdroj, jehož verzí popisovaný zdroj je
	Má verzí (hasVersion)	Odkaz na zdroj, který je verzí popisovaného zdroje
	Je nahrazen (isReplacedBy)	Odkaz na následný zdroj
	Nahrazuje (replaced)	Odkaz na zdroj, který popisovaný zdroj nahrazuje
	Související zdroj (isRequiredBy)	Odkaz na zdroj, který fyzicky nebo logicky vyžaduje také použití popisovaného zdroje
	Požaduje (requires)	Odkaz na zdroj, který je fyzicky nebo logicky vyžadován popisovaným zdrojem k použití
	Je částí (isPartOf)	Popisovaný zdroj je fyzickou nebo logickou součástí odkazovaného zdroje
	Má část (hasPart)	Popisovaný zdroj má jako jednu část odkazovaný zdroj
	Je citován kým (isReferencedBy)	Popisovaný zdroj je citován odkazovaným zdrojem
	Cituje (references)	Popisovaný zdroj cituje nebo jinak odkazuje na uvedené zdroje
	Jiný formát čeho (isFormatOf)	Popisovaný zdroj vyjadřuje stejný intelektuální obsah jako uvedený zdroj, ale v jiném formátu
	Má formát (hasFormat)	Odkaz na zdroj, který představuje stejný intelektuální obsah jako popisovaný zdroj, ale v jiném následně vytvořeném formátu
	Je zdrojem pro (isBasisFor)	Popisovaný zdroj sloužil jako předloha či inspirace pro odkazovaný zdroj
	Je založen na (isBasedOn)	Popisovaný zdroj byl předlohou či inspirací pro odkazovaný zdroj
Pokrytí (coverage)	Jurisdikce (jurisdiction)	Teritoriální nebo jiné označení právní platnosti zdroje
	Geografické (spatial)	Geografické pokrytí intelektuálního obsahu zdroje
	Časové (temporal)	Časové pokrytí intelektuálního obsahu zdroje
	Pomocí PSČ (postcode)	Pokrytí určené pomocí australských poštovních kódů
Mandát (mandate)	Nařízení (act)	Odkaz na státní nebo federální nařízení, na jehož základě zdroj vznikl
	Směrnice (regulation)	Odkaz na směrnici, která si vyžádala vznik zdroje
	Právní proces, kauza (case)	Odkaz na právní proces, na jehož základě zdroj vznikl
Obchodní význam zdroje pro organizaci (function)		
Uživatelské určení (audience)		

Spolu s formátem AGLS je možné používat libovolné řízené slovníky, třídění a tezaury. Kromě toho vytvořil Australský národní archiv některé vlastní doporučené řízené rejstříky pro vybrané prvky AGLS.

Pro prvek *funkce* se doporučuje využití speciálního Interaktivního funkčního tezauru australské vlády (AGIFT, Australian Government Interactive Functions Thesaurus), který byl vytvořen tak, aby běžnému uživateli usnadnil hledání ve vládních zdrojích také podle vládních úrovní. Zahrnuje všechny funkce, které jsou vykonávány ve třech vládních úrovních.

Jde ale o obecný tezaurus, který má zjednodušit vyhledávání uživatelům a pomoci jim orientovat se v základních funkcích a pojmech administrativy. Proto Australský národní archiv doporučuje kombinovat výrazy z tohoto obecného tezauru s tezaury podrobnějšími, které si vybraly nebo vytvořily jednotlivé instituce (Cunningham, 2003).

Zajímavé údaje nabízí řízený slovník pro prvek *uživatelské určení*. Kromě standardních možností dělení do kategorií (např. děti, mladiství do 16 let, rodiče, muži, zaměstnanci, zaměstnavatelé apod.) se v něm nachází řada specifických kategorií pro skupiny osob jako např. původní obyvatelé, nezaměstnaní, homosexuálové, vládní organizace a asociace, migranti, senioři, turisté apod. (Wilson, 2002a).

AGLS slouží i k popisu fyzických dokumentů. U nich je velice důležité schéma pro *dostupnost*. Nese v sobě především kontaktní informace na osobu či instituci, u níž je možné dokument požadovat, dále informace o geografické lokaci zdroje či služby a informace o ceně za zpřístupnění (Wilson, 2002c).

Pro prvky *autor*, *vydavatel* a *příspěvatel* se užívá také schéma AGLS Agent. Popisuje osoby a organizace. Schéma zahrnuje charakteristiky jako jména osob a organizací, kontaktní údaje a údaj, zda je subjekt z vládního sektoru či nikoliv (Wilson, 2002b).

Pro prvky *autor*, *vydavatel*, *příspěvatel*, *pokrytí* a *dostupnost* je určen řízený slovník AGLS jurisdiction. Tento slovník obsahuje termíny pro administrativní správní jednotky Austrálie, např. Jižní Austrálie, Tasmánie,... (Wilson, 2002e).

Pro kvalifikátor *dokument* v prvku *typ* je vhodné používat řízený slovník AGLS-document. Specifikuje typ dokumentu ve vládní sféře. Jde např. o formulář, stanovení hospodářské politiky, úřední oznámení, ale také o obecnější specifikaci typu domovská stránka, video apod. (Wilson, 2002d).

Opět u prvku *typ* ale s kvalifikátorem *typ služby* se používá jiný řízený slovník AGLS-service. Typem služby může být např. přihláška, výměna dat, finanční služby, granty, právní poradenství aj. (Wilson, 2002f).

Kromě ručního vytváření záznamu autorem lze využít také automatického vytvoření záznamu pomocí softwarové aplikace. K tomuto účelu se dají použít systémy pro management záznamů nebo balíčky pro publikování na webu, ale existují také speciální nástroje pro generování metadat ze zdrojů.

AGLS dovoluje vnořit metadata do popisovaného zdroje i vytvořit metadatový archiv externích záznamů. Externí záznamy musí být provázány s popisovanými zdroji. AGLS popisuje stejně jako Dublin Core pouze sémantiku metadatových prvků. Pro syntaktické účely se používá formátů HTML, XML a RDF (Cunningham, 1998). Při zápisu metadat AGLS do HTML nerozlišuje systém velká a malá písmena v prvcích. Jiná je ovšem situace u formátů XML a RDF. V případě užití těchto formátů systém velká a malá písmena rozlišuje.

Příklad 5: Záznam ve formátu AGLS na úrovni sbírky

```
<link rel="schema.AGLS" href="http://www.naa.gov.au/recordkeeping/gov_online/agls/1.2" >
<META NAME="DC.Identifier" SCHEME="URI"
CONTENT="http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html" >
<META NAME="DC.Title" CONTENT="Australian Government Locator Service (AGLS)" >
<META NAME="DC.Creator" SCHEME="GOLD" CONTENT="c=AU; co=Commonwealth of Australia; ou=National
Archives of Australia" >
<META NAME="DC.Subject" SCHEME="APAI" CONTENT="Archives; Information management;information retrieval" >
<META NAME="DC.Description" CONTENT="Provides access to information and resources about the Australian
Government Locator Service (AGLS) metadata standard, and how to implement it." >
<META NAME="DC.Publisher" CONTENT="corporateName=National Archives of Australia (NAA)" >
<META NAME="DC.Date.created" SCHEME="ISO8601" CONTENT="2000-03-31" >
<META NAME="DC.Type.aggregationLevel" CONTENT="collection" >
<META NAME="DC.Type.documentType" SCHEME="agls-document" CONTENT="promotional" >
<META NAME="DC.Format" SCHEME="IMT" CONTENT="text/html" >
<META NAME="DC.Language" SCHEME="RFC3066" CONTENT="en" >
<META NAME="DC.Coverage.jurisdiction" CONTENT="Commonwealth of Australia" >
<META NAME="DC.Rights" CONTENT="Copyright Commonwealth of Australia 2000" >
<META NAME="AGLS.Function" SCHEME="AGIFT" CONTENT="Recordkeeping standards" >
<META NAME="DC.Relation.references" SCHEME="URI" CONTENT="http://www.dca.gov.au/ogo/imsc/imscrypt.htm" >
<META NAME="AGLS.Availability" CONTENT="email=agls@naa.gov.au" >
```

AGLS umožňuje organizaci zvolit si úroveň popisu. Je možné provádět popis na úrovni sbírky i na úrovni konkrétní jednotky. Většina organizací dává přednost popisu na úrovni sbírky. Ukázkou záznamu na úrovni sbírky obsahuje příklad 5 (National Archives of Australia, 2002c). Úroveň popisu je v záznamu určena prvkem

Dublin Core „DC.Type.aggregationLevel“. Prvky, které jsou přidány k základním prvkům Dublin Core, jsou odlišeny pomocí prefixu „AGLS“.

Následující ukázka v příkladu 6 (National Archives of Australia, 2002c) obsahuje metadatový záznam dokumentu ze stránek Child Support Agency (<http://www.csa.gov.au/>). Popis se týká webové stránky se seznamem zdrojů ve formátech .html, .pdf nebo .rtf pro rodiče, zaměstnavatele a další subjekty, které mohou ovlivňovat proces péče o děti. Jednotlivé zdroje samostatně popsány nejsou, nebo alespoň není možné popis k nim zobrazit. Jde o kvalifikovaný popis, který kromě prvků Dublin Core používá doplňující prvky formátu AGLS. Hojně jsou použity odkazy na existující schémata.

Příklad 6: Záznam ve formátu AGLS popisující webovou stránku

```
<link rel="AGLS_Metadata_Element_Set_v1-3" href="http://www.agls.gov.au/AGLS_Metadata_Element_Set_v1-3.html" />
<meta name="DC.Creator" scheme="AglAgent" content="corporateName=Child Support Agency - Australia;
jurisdiction=Commonwealth of Australia; sector=Government" />
<meta name="DC.Publisher" scheme="AglAgent" content="corporateName=Child Support Agency - Australia;
jurisdiction=Commonwealth of Australia; sector=Government" />
<meta name="DC.Format" scheme="IMT" content="text/html" />
<meta name="DC.Rights" content="Copyright Commonwealth of Australia 2006" />
<meta name="AGLS.Function" scheme="AGIFT" content="Community services" />
<meta name="DC.Subject" scheme="TAGS" content="Child support" />
<meta name="DC.Type.aggregationLevel" content="collection" />
<meta name="DC.Type.documentType" scheme="agls-document" content="promotional" />
<meta name="DC.Description" content="List of CSA's corporate publications, and publications for parents, employers, legal
and community. Including child support handbook, financial help, parenting, mediation, family tax benefit, payment options,
overview for employers, newsletter for legal professions, repartnering resource and information kit for community
organisations, CSA's business plan, fact & figures reports, and FaCS annual report." />
<meta name="Description" content="List of CSA's corporate publications, and publications for parents, employers, legal
and community. Including child support handbook, financial help, parenting, mediation, family tax benefit, payment options,
overview for employers, newsletter for legal professions, repartnering resource and information kit for community
organisations, CSA's business plan, fact & figures reports, and FaCS annual report." />
<meta name="DC.Date.created" scheme="W3CDTF" content="2005-08-29" />
<meta name="DC.Date.modified" scheme="W3CDTF" content="2006-08-08" />
<meta name="DC.Identifier" scheme="URI" content="http://www.csa.gov.au/publications/index.php" />
<meta name="DC.Title" content="Publications" />
<meta http-equiv="Keywords" content="handbook, financial, security, mediation, separating, pay, direct, payment, options,
Business, Agency Agreement, Workplace Diversity, Facts and Figures, Multicultural, Australia and Comparisons, plans,
reports" />
```

4.2.3 Shrnutí informací k formátu AGLS

AGLS je formát, který je podporován australskou vládou. Na jeho vývoji a implementaci se podílí řada významných australských organizací. Úspěchem je, myslím, i to, že i když AGLS vychází z formátu Dublin Core, stal se sám základem pro formáty jiné. Velkou zásluhu na tom jistě má i snaha australské vlády používat tento formát pro vládní a úřední dokumenty a zdroje.

4.3 Education Network Australia (EdNA)



Zajímavou službou, která byla vyvinuta také v Austrálii a používá k popisu informačních zdrojů upravený formát Dublin Core je Síť pro vzdělávání v Austrálii EdNA (Education Network Australia, <http://www.edna.edu.au/>). EdNA Online je informační služba, která přispívá ke zvýšení využití sítě Internet v oblasti vzdělávání, výuky a školství v Austrálii. Službu EdNA finančně podporuje australská vláda. Nabízí zdarma služby pro australské učitele a lektory. Databáze záznamů webových zdrojů pro vzdělávání a školství a adresář pro vzdělávání a školství v Austrálii jsou dvěma základními službami. Cílem je, aby na službě EdNA spolupracovala celá vzdělávací komunita ze všech částí Austrálie. Na stránkách věnovaných této službě může každý navrhnout konkrétní webovou stránku, jejíž záznam by měl být zařazen do databáze záznamů o zdrojích. Je zde také sekce pro potenciální nové členy, kteří by chtěli na službě participovat (EdNA, 2007a).

Vznik původního projektu EdNA iniciovala v roce 1995 vláda Australského svazu s vědomím velkého významu vzdělání pro společnost. Již od začátku byl projekt koncipován jako kooperativní projekt všech vládních vzdělávacích složek v Austrálii. Na jeho financování se podílí vláda celého Australského svazu, zapojeno je všech šest států a vlády dvou teritorií. Na vytváření databáze se v rámci distribuovaného modelu podílí řada autorit australského vzdělávání, a to jak vládních tak nevládních. První aktivitou těchto organizací byla snaha vytvořit rozsáhlou kooperativní celonárodní síť vzdělávacích institucí, zlepšit vzájemné kontakty a informovanost mezi těmito institucemi. Následovalo založení databáze EdNA, do které začaly být ukládány záznamy kvalitních webových zdrojů pro vzdělávání.

Správou společně vytvářené databáze EdNA byla pověřena nezisková společnost education.au limited, která byla pro tyto účely založena v červenci 1996. Představenstvo společnosti education.au limited tvoří zástupci hlavních investorů služby EdNA a předseda je vždy jmenován vládou Australského svazu. V následujících letech se společnost education.au limited ujala správy dalších služeb a funkcí portálu, který na základě projektu EdNA vznikal.

Databáze, nad níž je zajištěno vyhledávání na portálu EdNA, obsahuje záznamy webových zdrojů spolupracujících australských vzdělávacích institucí,

ať již z vládního nebo nevládního sektoru. Základem databáze jsou metadatové záznamy zdrojů, které pečlivě vybírají podle stanovených kritérií odborníci z oblasti vzdělávání nebo pracovníci společnosti education.au limited. Záznamy jsou dále zpracovávány, propojovány s dalšími záznamy a indexovány. Následně je možné vyhledat je jak pomocí vyhledávání, tak pomocí listování v kategoriích. Kromě manuálně zpracovávaných a manuálně zařazovaných záznamů se do databáze dostávají také záznamy sklizené harvesterem. Sklizení probíhá na základě odkazů ve zdrojích již zařazených záznamů a na základě dalších pravidel stanovených společností education.au limited (White, 2005).

Portál EdNA Online se průběžně vyvíjí. Několikrát byl upraven vzhled portálu a postupně jsou přidávány další služby. Jednou z nich je služba MyEdNA, která dovoluje vytvořit si vlastní EdNA konto a v něm přednastavit údaje, které se mají na uživatelské stránce zobrazovat. Je pak možné přihlásit se ke službě MyEdNA a okamžitě mít k dispozici vyhledávací formuláře, aktuality z oblasti vzdělávání, aktuálně pořádané konference a jiné odborné akce nejen v Austrálii ale i celosvětový přehled, nebo zobrazené novinky z vybraných elektronických časopisů nebo portálů organizací. Pro definici vzhledu se používá rozhraní s jednoduchou navigací.

Dostupná je také služba RSS, která zajistí včasné informování o novinkách portálu. Plně využit byl také formát XML a metadata. S pomocí nástrojů webových služeb a právě metadat je možné distribuovat vzdělávacím institucím materiály zaměřené na oblast vzdělávání, ve které působí (EdNA, 2003b). EdNA nabízí také zasílání aktuálních novinek z oblasti vzdělávání na přenosné kapesní počítače.

Pro zajištění bezproblémové spolupráce poskytuje portál EdNA vzdělávacím institucím nástroje a standardy pro správu informací a informační systémy, spravuje metadatový repozitář a nad ním strukturu kategorií a další nástroje pro vyhledávání. Portál EdNA dosahuje svojí komplexností díky široké základně australských organizací, které jej financují a které se na jeho fungování podílí. Spolupráce je zajištěna také s jednotlivými odborníky z oblasti vzdělávání a výchovy prostřednictvím zájmových skupin, online diskusí a elektronických konferencí spravovaných portálem EdNA, kterých je více než čtyři sta. Cílovou skupinou všech služeb jsou odborníci z oblasti vzdělávání i různé vzdělávací instituce a organizace a v neposlední řadě samotní studenti a žáci (White, 2003).

4.3.1 Popisné jednotky

Většina zdrojů je popisována na úrovni jednotky. Záznamy zdrojů, které jsou do databáze EdNA zařazeny, musí splňovat několik kritérií. Především musí obsahově odpovídat australskému vzdělávacímu systému, musí být obsahově správné a aktuální, musí být určena odpovědnost, nesmí popírat australské právo, lidská práva, nebo porušovat autorská práva. Zdroje by měly samozřejmě odpovídat svým stylem určené cílové skupině a účelu.

Ve svých doporučeních pro spolupracující organizace upozorňuje EdNA na úskalí popisu na úrovni sbírky. Přestože tuto možnost nevylučuje a uvědomuje si, že pro řadu institucí je takovýto popis jednodušší, varuje před skutečností, že pak jednotlivé webové stránky a elektronické zdroje nebudou vyhledatelné. Naopak v případě aplikace metadatového popisu pro celou sbírku na úroveň jednotlivých stránek nebo zdrojů může dojít při vyhledávání k zahlcení uživatele řadou nerelevantních zdrojů. Z toho je zřejmé, že se EdNA snaží své partnery motivovat k popisu zdrojů na úrovni jednotky.

Záznamy zdrojů jsou před zařazením do databáze pečlivě indexovány a propojovány se souvisejícími zdroji. Vzniká tak základní klíčová databáze, ve které jsou záznamy zdrojů vyhledatelné také přes kategorie, do kterých jsou zařazené, a kromě toho širší databáze, ve které je možné vyhledat jak základní tak související zdroje pomocí vyhledávání přes klíčová slova (White, 2005).

4.3.2 Metadatové záznamy

Metadata, která služba EdNA využívá, jsou založena na formátu Dublin Core a jsou v souladu také s formátem AGLS. Původní verze metadatového formátu pro projekt EdNA byla publikována v roce 1998. Současná verze 1.1 je z roku 2000. Základem metadatového formátu je formát Dublin Core se svými patnácti prvky. Používají se také kvalifikátory těchto prvků a schémata zápisu. Upřednostňována jsou samozřejmě australská schémata a klasifikace. Při popisu je kladen důraz na zachování rovnováhy mezi popisem příliš obecným a naopak příliš podrobným. Cílem je snadná výměna záznamů v různých formátech, především záznamů

ve formátech Dublin Core, AGLS a IMS⁷. Prvky Dublin Core, tak jak jsou používány portálem EdNA, jsou uvedeny v příloze 2 (EdNA, 2002).

Kromě prvků formátu Dublin Core je definováno dalších osm prvků specifických pro formát EdNA. Jde především o prvky, které souvisí s určením cílových uživatelů, dále o prvky související se zpracováním a hodnocením zdrojů a také administrativní prvky, které mají napomoci snadnější správě a údržbě databáze EdNA. Tyto prvky jsou uvedeny v tabulce 4.

Tab. 4: Vlastní metadatové prvky formátu EdNA

Prvek	Kvalifikátor	Definice	Schéma zápisu
EDNA.Audience		Uživatelská kategorie, pro kterou je určen obsah zdroje. Doporučení použít některý z řízených slovníků EdNA	edna-audience: student teacher/lecturer parent/carer administrator community member
	edna-sector	Uživatelská komunita, pro který je určen obsah zdroje	edna-sector: preschool school VET higher education ACE
	edna-userlevel	Stupeň pokročilosti uživatelů, pro kterou je určen obsah zdroje. Stupeň určuje počet let, kdy probíhalo vzdělávání uživatelů. Nultý stupeň reprezentuje předškolní výchovu	edna-userlevel: Stupně 0-13
EDNA.Approver		Email osoby nebo organizace, která schválila zařazení zdroje do databáze EdNA	
EDNA.CategoryCode		Číselný kód předmětové kategorie EdNA	
EDNA.Entered		Administrativní údaje o zařazení datové jednotky do databáze pro správu databáze	
EDNA.Indexing		Určení rozsahu, v jakém mají být indexovány zdroje odkazované v popisovaném zdroji	
EDNA.Review		Hodnocení zdroje třetí stranou. Může zahrnovat též pouze odkaz na obsáhlejší hodnocení. Doporučuje se použít spolu s prvkem EDNA.Reviewer	
EDNA.Reviewer		Jméno osoby, organizace nebo jiné autority odpovědné za hodnocení zdroje	
EDNA.Version		Použitá verze metadatového formátu EdNA	

⁷ Metadatové schéma IMS je zaměřené na vzdělávání a výchovu, spravuje ho mezinárodní nezisková organizace IMS Global Learning Consortium.

Záznamy v databázi EdNA popisují objekty většinou na úrovni jednotlivého zdroje. Jsou vytvářeny ručně. Do celé databáze jsou ale data také sklízena přes protokol OAI-PMH. Jde o záznamy webových stránek a elektronických zdrojů zpřístupňovaných vybranými vzdělávacími institucemi a organizacemi, které své zdroje obvykle také vytváří ručně. Záznamy jsou následně zpřístupněny přes jednotné vyhledávání na portálu EdNA. Výměna a sdílení metadatových záznamů je jedním z cílů portálu EdNA. EdNA nabízí ostatním institucím také svoje metadatové záznamy (EdNA, 2007d).

EdNA doporučuje svým partnerům používání číselníků, rejstříků a tezaurů pro vyplňování obsahu některých prvků. Pro různé oblasti vzdělávání jsou doporučovány různé tezaury. Např. pro oblast vysokoškolského vzdělávání je doporučen *Australský tezaurus pro deskriptory z oblasti vzdělávání* (ATED)⁸, pro oblast profesního vzdělávání a výukové kurzy tezaurus *Profesní vzdělávání a školení* (VOCED)⁹ nebo *Australský seznam povolání* (OZJAC)¹⁰.

Pokud je ve sklizených metadatových záznamech obsažen kód předmětové kategorie EdNA, je zdroj navíc zařazen pod danou kategorii a následně je vyhledatelný i přes hierarchický strom těchto kategorií a nikoliv pouze přes hledání. EdNA definuje přes dva tisíce číselných kódů pro své kategorie.

EdNA nechává plně na vnitřním rozhodnutí jednotlivých spolupracujících institucí, v jaké fázi vzniku a zpřístupnění zdroje bude metadatový záznam vytvářet a zda budou metadata vnořena nebo uložena v externí databázi. Pro vytváření metadatových záznamů nabízí pomocné nástroje.

Má-li být záznam sklizen, musí obsahovat prvky DC.Identifier, DC.Title, DC.Description a DC.Subject nebo AGLS.function. Záznamy obsahující více prvků mohou být sklizeny a to i v případě, že obsahují prvky jiných metadatových formátů, než jaké používá databáze EdNA (EdNA, 2006). Příklad 7 obsahuje záznam Ministerstva školství Tasmánie, který obsahuje jak prvky formátu Dublin Core, tak prvky AGLS a EdNA. Záznamy jsou navíc obohaceny vlastními prvky TASEOS (EdNA, 2003a).

⁸ Australian Thesaurus of Education Descriptors. Bližší informace na WWW: <http://www.acer.edu.au/library/Catalogues/ated/thesaurus.html>

⁹ Vocational Education & Training. Bližší informace na WWW: <http://www.voced.edu.au/thes.htm>

¹⁰ Australian List of Occupations. Bližší informace na WWW: <http://www.curriculum.edu.au/ozjacweb/>

Příklad 7: Metadatový záznam ve formátu EdNA s použitím prvků z jiných formátů

```
<meta name="DESCRIPTION" content="The homepage for Discover, a web site promoting participation and contribution, encouraging the sharing of ideas, practices and resources developed by the Department of Education, Tasmania">
<meta name="KEYWORDS" content="teachers, parents, netlearners, students, community, department, databank, collaborative, Tasmania, education">
<!--Dublin Core Metadata-->
<meta name="DC.Title" content="Discover (Home)">
<meta name="DC.Creator" content="Department of Education">
<meta name="DC.Subject" content="teachers, parents, netlearners, students, community, department, databank, collaborative, Tasmania, education">
<meta name="DC.Description" content="The homepage for Discover, a web site promoting participation and contribution, encouraging the sharing of ideas, practices and resources developed by the Department of Education, Tasmania">
<meta name="DC.Publisher" content="Department of Education, Tasmania">
<meta name="DC.Contributor" content=" " >
<meta name="DC.Date" content="1999-06-28">
<meta name="DC.Type" content="text">
<meta name="DC.Format" content="text/html">
<meta name="DC.Identifier" content="http://www.discover.tased.edu.au/default.htm">
<meta name="DC.Source" content=" " >
<meta name="DC.Language" content="en">
<meta name="DC.Relation" content=" " >
<meta name="DC.Coverage" content="TAS - Tasmania">
<meta name="DC.Rights" content="http://www.discover.tased.edu.au/webgov/rights.htm">
<!--AGLS metadata-->
<meta name="AGLS.Function" content=" " >
<!--EDNA metadata-->
<meta name="EDNA.Userlevel" content="teacher">
<meta name="EDNA.Categories" content="4710">
<meta name="EDNA.Indexing" content="2">
<!--TASEOS metadata-->
<meta name="TASEOS.CreatorName" content="Discover">
<meta name="TASEOS.CreatorEmail" content="discover.team@central.tased.edu.au">
<meta name="TASEOS.CustodianPersonalName" content="Elizabeth Brook">
<meta name="TASEOS.CustodianCorporateName" content="Web Publisher">
<meta name="TASEOS.CustodianEmail" content="elizabeth.brook@central.tased.edu.au">
<meta name="TASEOS.AuthorityPersonalName" content="David Hanlon">
<meta name="TASEOS.AuthorityCorporateName" content="Director">
<meta name="TASEOS.AuthorityEmail" content="David.Hanlon@central.tased.edu.au">
<meta name="TASEOS.ResourceID" content="Discover">
<meta name="TASEOS.ResourceStatus" content="live">
<meta name="TASEOS.DateModified LANG=" content="2000-11-30">
<meta name="TASEOS.ReviewFrequency" content="180">
<meta name="TASEOS.Sensitivity" content="low">
<meta name="TASEOS.Significance" content="normal">
<meta name="TASEOS.System" content="Discover">
<meta name="TASEOS.SiteClass" content="A">
<meta name="TASEOS.SiteType" content="internet">
```

Vzhledem ke kompatibilitě s formátem Dublin Core i AGLS není problém do databáze EdNA naimportovat nebo stáhnout přes protokol OAI-PMH data z jiných databází. I pro spolupracující organizace to znamená přínos, protože jejich data jsou snadněji vyhledatelná díky velkému využití služby EdNA Online. Záznamy obsahují i informace o vlastnících, takže vyhledaný zdroj lze také získat (EdNA, 2003a).

4.3.3 Vyhledávání

Kvalitně vytvořené metadatové záznamy hrají důležitou roli také při následném vyhledávání v databázi EdNA. Na základě údajů z metadatových záznamů jsou záznamy zařazována pod předmětové kategorie EdNA, které jsou také

jednou z možných cest při hledání odpovídajícího informačního zdroje (EdNA, 2002).

Vyhledávání probíhá nad základní databází, která obsahuje zhruba 20 000 kvalitních prověřených zdrojů. Kromě toho je možné použít distribuované hledání a vyhledávat nad 1,6 miliony zdrojů ze spolupracujících repozitářů (EdNA, 2007b).

Portál EdNA nabízí několik typů vyhledávání. Základní vyhledávací formulář nabízí pouze jedno vyhledávací pole bez možnosti rozlišení, v jakých údajích má vyhledávání probíhat.

Pokročilé vyhledávání dovoluje specifikovat, zda mají být hledané výrazy považovány za frázi, nebo zda mají být spojeny pomocí operátoru AND nebo OR. Dále je možné omezit hledání na vybranou předmětovou kategorii EdNA nebo na určitou službu portálu EdNA. Je také možné použít hledání pomocí výrazů z dostupných tezaurů.

Distribuované hledání dovoluje omezit vyhledávání na vybrané repozitáře.

Výsledky vyhledávání jsou primárně seřazeny podle relevance. V přehledu záznamů se zobrazuje titul zdroje. Nebylo-li hledání provedeno přes formulář pro metadatové vyhledávání je uvedena relevance zdroje vůči dotazu, kategorie EdNA a krátký popis zdroje. U každého záznamu je možné zobrazit metadatový záznam. Záznam se zobrazí v tabulkovém náhledu. Přes URL, které je v záznamu uvedeno, je možné přejít na zdroj samotný. Často jsou metadata vnořena přímo v hlavičce zdroje.

Záznam v příkladu 8 obsahuje záznam zdroje, který byl vyhledaný přes portál EdNA. Metadata jsou v tomto zdroji vnořena. V záznamu jsou přítomny prvky z formátů Dublin Core, AGLS i EdNA. Záznam je kvalifikovaný a obsahuje všechna povinná pole (DC.Identifier, DC.Title, DC.Description, oba alternativní prvky DC.Subject i AGLS.function). Hodnoty některých prvků jsou převzaty z doporučených schémat a klasifikací. V záznamu jsou vidět třídění a kódy používané přímo metadatovým formátem EdNA (EDNA.audience, EDNA.Categories) i Dublin Core (pro kódované vyjádření data norma ISO 8601).

Příklad 8: Ukázka metadatového záznamu z databáze EdNA

```
<META NAME="KEYWORDS" CONTENT="Schools, home, education, home school, registration, faq,">
<META NAME="DC.Title" CONTENT="ACT Department of Education and Training - Schools">
<META NAME="DC.Subject" SCHEME="ACT Government Thesaurus" CONTENT="Community Relations">
<META NAME="AGLS.FUNCTION" SCHEME="TVKAAA" CONTENT="Community Relations">
<META NAME="DC.Description" CONTENT="Information about schools in the ACT including links to their websites,
curriculum information, priority enrolment areas, and assessment and reporting.">
<META NAME="Description" CONTENT="Information about schools in the ACT including links to their websites,
curriculum information, priority enrolment areas, and assessment and reporting.">
<META NAME="DC.Creator" SCHEME="ACTIVERS" CONTENT="corporateName=ACT Department of Education
and Training ; SectionName=Education Branch; PositionTitle=">
<META NAME="DC.Identifier" CONTENT="http://www.det.act.gov.au/schools/schools.htm">
<META NAME="DC.Date.Created" SCHEME="ISO8601" CONTENT="2005-11-22">
<META NAME="DC.Date.Modified" SCHEME="ISO8601" CONTENT="2005-11-22">
<META NAME="DC.Date.Valid" SCHEME="DCMIPeriod" CONTENT="">
<META NAME="DC.Publisher" CONTENT="ACT Department of Education and Training">
<META NAME="EDNA.Audience" SCHEME="edna-sector" CONTENT="preschool, school, vocational education
and training">
<META NAME="EDNA.Audience" SCHEME="edna-audience" CONTENT="teacher/lecturer, parent/carers, student,
administrator, community member">
<META NAME="EDNA.Categories" CONTENT="4650">
<META NAME="EDNA.Version" CONTENT="1.1">
```

4.3.4 Shrnutí informací k portálu EdNA

Využití portálu EdNA je rozsáhlé. Lze to demonstrovat na průměrných počtech přístupů k portálu (přes 152 000 za měsíc), zvyšujícím se využití zasílání informací pomocí RSS, využití služby MyEdNA (přes 9 000 registrací), nebo na aktivní účasti více než 42 000 osob v elektronických konferencích, které EdNA provozuje (EdNA, 2007d). Portál EdNA je jednou z ukázek velmi dobře fungující koordinované spolupráce mezi řadou různých institucí v jednom oboru. Vzhledem k tomu, že i nadále má portál EdNA ambice se vyvíjet, rozšiřovat stávající a implementovat nové služby, lze očekávat jeho úspěšné fungování i v následujících letech.

4.4 Popis vysokoškolských prací na příkladu NDLTD

4.4.1 Archivace vysokoškolských prací

Snaha archivovat a zpřístupňovat disertační a diplomové práce v elektronické podobě je patrná v celé řadě zemí. Jednou z nejdéle působících institucí v této oblasti je americká společnost University Microfilms (dříve UMI, dnes součást společnosti ProQuest), která vytváří a na komerční bázi zpřístupňuje databázi doktorských disertačních prací z USA a Kanady. Nejstarší tituly v bázi jsou z roku 1861. Celkem je v databázi přes 2,2 miliony záznamů a zhruba 1,9 milionů těchto vysokoškolských prací je dostupných v plných textech. V současnosti je databáze nabízena jako jedna z databází centra ProQuest pod názvem ProQuest Dissertations & Theses (ProQuest, 2007).

Kromě čistě komerčních aktivit v této oblasti se také od konce 20. století projevuje snaha jednotlivých vysokých škol, vzdělávacích institucí a asociací a jejich konsorcií o archivaci a zpřístupňování obhájených prací svých studentů. Vzniká řada elektronických archivů a fondů, které jsou založené na dobrovolné spolupráci různých institucí z různých zemí (NDLTD), vytváří se také mnoho národních archivů na základě spolupráce řady vzdělávacích institucí v jedné zemi (např. Německo, Austrálie, Francie).

V České republice byla v dubnu 2004 založena *Komise pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací*, která se snaží o koordinaci a spolupráci vysokých škol při budování archivů vysokoškolských prací. V listopadu 2005 zveřejnila Komise návrh národního metadatového formátu pro popis vysokoškolských prací. Vycházela přitom ze znalostí a zkušeností zahraničních systémů pro popis vysokoškolských prací i z místních možností a poměrů. Komise eviduje snahy jednotlivých vysokých škol o budování archivů vysokoškolských prací a snaží se o sjednocení praxe (Asociace knihoven vysokých škol České republiky, 2004).

Archivovány a popisovány jsou obvykle disertační a habilitační práce, mohou být ale popisovány i práce diplomové, případně jiné vysokoškolské práce. Přestože bývají práce často složeny z více typů dat, bývají považovány za jeden celek a při zpracování se k nim přistupuje jako k jedné popisné jednotce.

Při archivování vysokoškolských prací a jejich zpřístupňování je nutné vyřešit nejen celou zpracovatelskou linku včetně uložení celé práce a metadatového záznamu o ní a technické zajištění celého procesu, ale také ošetřit autorská práva. Obvykle musí dát autor práce písemný souhlas s jejím zveřejněním (Bratková, 2003).

Tato práce nabídne ukázkou jedné zahraničních digitální knihovny zaměřené na vysokoškolské práce.

4.4.2 NDLTD



Na vytváření *Digitální knihovny diplomových a disertačních prací* (Networked Digital Library of Theses and Dissertations, NDLTD, <http://www.ndltd.org/>) se podílí mezinárodní společenství univerzit a dalších vzdělávacích organizací. Centrem této digitální knihovny je Virginský polytechnický institut a státní univerzita (Virginia Polytechnic Institute and State University), který je známý také pod zkratkou „Virginia Tech“.

Digitální knihovna NDLTD byla založena v roce 1997. Na počátku stál jednoletý výzkumný projekt zahájený v roce 1996, který vedl právě Virginský polytechnický institut a státní univerzita. Původně měl projekt ambice vybudovat národní digitální knihovnu diplomových a disertačních prací. Nicméně na tvorbě této digitální knihovny se nepodílí všechny univerzity a nedaří se do ní získávat všechny obhájené vysokoškolské práce v zemi (Bratková, 2002).

V současnosti patří mezi členy mezinárodní federace NDLTD řada vzdělávacích organizací z osmnácti zemí. Kromě institucí z USA na NDLTD participují instituce z Velké Británie, Německa, Austrálie, Koreje, Indie, Kanady, Mexika, Číny, Ruska a řady dalších zemí. Pro příklad jmenujme alespoň Australskou národní univerzitu (Australian National University), Čínskou univerzitu v Hong Kongu (Chinese University of Hong Kong), Městskou univerzitu v Londýně (City University, London), Humboldtovu univerzitu v Berlíně (Humboldt-Universität zu Berlin), Univerzitu v Lundu (Lund University, Sweden) a další (Scheid, 2000; NDLTD, 2004b).

Projekt digitální knihovny NDLTD vycházel z mnoha jednání a zkušeností z jiných projektů. Úvahy o budování sbírky elektronických vysokoškolských prací se objevovaly už od roku 1987, kdy se konalo setkání v Ann Arboru v Michiganu.

Organizátorem tohoto setkání byla UMI (University Microfilms). Následovalo několik projektů, kterých se účastnili také zástupci Virginia Tech. Jedním z těchto projektů byl také projekt Elektronická knihovna Monticello (Monticello Electronic Library). Část tohoto projektu byla zaměřena i na vysokoškolské práce. Za finanční podpory Výzkumné asociace jihovýchodních univerzit (Southeastern Universities Research Association, SURA) a Sítě jihovýchodních knihoven (Southeastern Library Network, Solinet) byl vyvinut software, který umožnil studentům publikovat své vysokoškolské práce ve formátech SGML nebo PDF. V rámci projektu byly připraveny podklady a software pro vytvoření metadatového formátu pro elektronické diplomové a disertační práce ETD (Electronic Theses and Dissertations) (NDLTD, 2004a). Dvě výše uvedené organizace sponzorovaly další výzkumy a vývoj v oblasti archivace elektronických vysokoškolských prací. Na sponzorování pilotní fáze projektu NDLTD od roku 1996 se podílelo americké ministerstvo školství (NDLTD, 2001).

Důležitou otázkou, která byla v projektu NDLTD řešena, byl celý proces související s přípravou, tvorbou i následným archivováním a zpřístupňováním vysokoškolských prací. Proces byl řešen již od stadia přípravy studenta na tvorbu vysokoškolské práce, přes poskytnutí nástrojů pro vypracování práce v elektronické formě s možností použití různých dříve nestandardních prvků (video, zvuk,...). Po dokončení práce bylo doporučeno převést ji do formátu PDF. Nebyly opomenuty ani otázky předložení a obhájení práce a samozřejmě ani vytváření metadatového záznamu a uložení práce spolu se záznamem do lokálního katalogu a také do databáze NDLTD. Zpřístupňování elektronických vysokoškolských prací představovalo a stále představuje problém kvůli autorským právům. Stává se, že autoři kladou podmínky na alespoň částečné omezení přístupu k práci (Bratková, 2002).

NDLTD vychází i z předpokladu, že ukládání elektronických vysokoškolských prací šetří knihovnám finanční náklady. Sice se nesnižují náklady na katalogizaci prací, ale odpadají náklady na vázání prací, jejich zabezpečování, polepování štítky, manipulaci se zdroji a jejich udržování a výpůjční služby. K práci může přistupovat více uživatelů najednou.

NDLTD se snaží napomáhat zvyšování kvality výuky, sdílení znalostí a výsledků akademických výzkumů, pomáhá univerzitám budovat informační

infrastrukturu a rozšiřovat a využívat digitální knihovny. Studenty se NDLTD snaží přesvědčit, aby zpřístupňovali své práce v elektronické podobě a umožnili tak jejich další využití (Scheid, 2000). Pro studenty je jednodušší vytvářet práce v elektronické podobě. Mohou uplatnit více kreativity, využít nejen hypertextové odkazy, ale kombinovat v práci text s grafickými soubory, zvukovými ukázkami, mohou vytvořit multimediální materiál. Institucím i studentům poskytuje NDLTD softwarové nástroje a informační podporu. Kromě informačních letáků, metodik a postupů je zpřístupňován různý software pro vytváření vysokoškolských prací v elektronickém formátu nebo pro vytváření metadatových záznamů

4.4.2.1 *Metadatové záznamy*

Popisnými jednotkami jsou vysokoškolské práce, především diplomové a disertační. Používá se pro ně upravený metadatový formát Dublin Core označovaný EDT-MS. Formát využívá třináct prvků Dublin Core a doplňuje prvek *thesis.degree* pro vysokoškolské práce. Ten je používán spolu s kvalifikátory. Z formátu Dublin Core nevyužívá EDT-MS prvky zdroj (source) a vztah (relation). Tabulka 5 obsahuje výčet prvků EDT-MS s kvalifikátory.

Záznam vysokoškolské práce ve formátu EDT-MS ukazuje příklad 9. Jsou v něm standardní údaje formátu DC a ke konci záznamu informace specifické pro vysokoškolské práce (titul, obor, univerzita, která titul udělila) (NDLTD, 2007b). Zajímavé je, že prvky nejsou uvozeny žádným prefixem a informace o souvisejícím metadatovém formátu jsou uvedeny pouze v úvodním prvku <thesis> (NDLTD, 2007b). Nepodařilo se mi zjistit, zda je to běžné u všech záznamů v databázi NDLTD.

Každá spolupracující organizace má přesto možnost zvolit si jako interní formát pro vlastní databázi jakýkoliv formát, nebo EDT-MS upravit podle svých potřeb. Popis by měl ale vždy zachycovat autora, název a kontext práce (NDLTD, 2007b). EDT-MS slouží jako společný formát pro konverzi z jednoho lokálního formátu do formátu jiné instituce. Aby bylo sdílení a výměna záznamů jednodušší, vytvořila NDLTD konverzní tabulku pro formáty EDT-MS a MARC21. Tabulka je k nahlédnutí v příloze 3.

Tab. 5: Prvky EDT-MS

Prvek NDLTD	Kvalifikátor	Popis	Poznámky
dc.title (název)		Název vysokoškolské práce, který je uveden na titulní straně	Povinný, opakovatelný
	alternative (alternativní název)	alternativní název diplomové práce či disertace	Volitelný, Opakovatelný
dc.creator (tvůrce)		Primární odpovědnost za obsah vysokoškolské práce. Jméno se uvádí ve formě volného textu v podobě, v jaké je uvedeno na titulní straně. Může být učiněn odkaz na autoritní záznam, pokud je dostupný	Povinný, Opakovatelný
dc.subject (předmět)		Téma vysokoškolské práce. Opět se doporučuje užívat klíčová slova a deskriptory, které jsou uvedeny v disertační či diplomové práci. U tohoto pole je možné využít řízené slovníky jednotlivých institucí	Povinný, opakovatelný
dc.description (popis)		V případě vysokoškolských prací je v tomto prvku uváděn abstrakt, který má být součástí těchto prací. tento prvek může být také specifikován pomocí kvalifikátoru <dc.description.abstract>	Volitelný, opakovatelný
	abstract (popis - abstrakt)	Plný text abstraktu vysokoškolské práce	Volitelný, opakovatelný
	note (popis - poznámka)	Doplňující informace týkající se diplomové či disertační práce	Volitelný, opakovatelný
	release (popis - oznámení)	Popis verze práce. Užívá se pro opravy tiskových chyb	Volitelný, opakovatelný
dc.publisher (vydavatel)		Osoba či instituce, která je odpovědná za zveřejnění vysokoškolské práce. An entity responsible for making the resource available. Pole může být vyplněno volným textem, nebo mohou být použity soubory autorit	Volitelný, opakovatelný
dc.contributor (příspěvatele)		Osoba zodpovědná za některé části práce. Může jít o spoluautory či školitele nebo oponenta	Volitelný, opakovatelný
	role (role příspěvatele)	Role, kterou přispěl spoluautor ke vzniku vysokoškolské práce (např. spoluautor, konzultant, oponent, ...)	Volitelný
dc.date (datum)		Datum vzniku práce. Obvykle je v tomto prvku uváděno datum z titulní stránky práce	Povinný
dc.type (typ)		Typ zdroje	Povinný opakovatelný
dc.format (formát)		Fyzické či digitální vyjádření zdroje. V případě elektronických vysokoškolských prací jde o uvedení formátů, ve kterých je práce dostupná	Volitelný, opakovatelný
dc.identifier (identifikátor)		Jedinečný identifikátor zdroje	Povinný, opakovatelný
dc.language (jazyk)		Jazyk textu práce	Volitelný, opakovatelný
dc.coverage (pokrytí)		Tematické pokrytí obsahu zdroje	Volitelný, opakovatelný
dc.rights (autorská práva)		Informace o autorských právech, která jsou spjata se zdrojem. Obvykle jde o podmínky přístupu ke zdroji	Volitelný, opakovatelný
thesis.degree		Speciální prvek pro vysokoškolské práce. Používá se s doplňujícími kvalifikátory	
	name (jméno hodnotí)	Hodnota, která je spojena s daným typem vysokoškolské práce	Volitelný, opakovatelný
	level (úroveň hodnotí)	Dosažený stupeň vzdělání, který je spjat s danou vysokoškolskou prací	Volitelný, opakovatelný
	discipline (obor hodnotí)	Oblast lidského poznání, které se týká obsah práce. Obvykle jde o jméno fakulty, kde byla práce vytvořena	Volitelný, opakovatelný
	grantor (garant hodnotí)	Instituce, která garantuje dosaženou úroveň vzdělání a dosažený titul, které jsou spojeny s obhájenou prací. Opět může být obsah uveden volným textem, nebo vybrán ze souboru autorit	Volitelný, opakovatelný

Příklad 9: Ukázka záznamu elektronické vysokoškolské práce ve formátu EDT-MS

```
<thesis xmlns="http://www.ndltd.org/standards/metadata/etdms/1.0/"
      xsi:schemaLocation="http://www.ndltd.org/standards/metadata/etdms/1.0/
      http://www.ndltd.org/standards/metadata/etdms/1.0/etdms.xsd">
  <title>Conceptual Development and Empirical Testing of an Outdoor Recreation Experience Model: The Recreation
  Experience Matrix (REM)</title>
  <creator>Walker, Gordon James</creator>
  <subject>outdoor recreation</subject>
  <subject>recreation experience preference scales</subject>
  <subject>recreation experience matrix</subject>
  <subject>recreation opportunity spectrum</subject>
  <description>This dissertation examines four issues, including:
    (a) whether outdoor recreation experiences not included in the Recreation Experience Preference (REP) scales exist; (b)
    whether these experiences can be categorized using a framework called the Recreation Experience Matrix (REM); (c) how
    well the Recreation Opportunity Spectrum (ROS) variables of activity, setting, and expertise explain the types of experiences
    outdoor recreationists receive; and (d) how well two new variables--primary mode and mode dependence--explain the types
    of experiences outdoor recreationists receive. In order to address these issues, an on-site questionnaire was distributed at
    Mount Rogers National Recreation Area in Virginia during October and November, 1995. a total of 410 people completed
    this questionnaire. Of these, 336 provided useable addresses for a follow-up mail-out questionnaire, with 169 (50.3%)
    actually returning it. After performing a variety of statistical analyses, it was found that: (a) some outdoor recreationists did
    report having non-REP experiences involving identity, cognition, absorption, and self-concept; (b) indirect support does exist
    for classifying outdoor recreation experiences using the REM framework; and (c) the ROS variables of activity, setting,
    and expertise, do explain some outdoor recreation experiences, as do the new variables of primary mode and mode
    dependence.
  </description>
  <publisher>Virginia Polytechnic Institute and State University</publisher>
  <contributor role="committee_member">Daniel R. Williams</contributor>
  <contributor role="committee_member">K. Jill Kiecolt</contributor>
  <contributor role="committee_member">Bradley R. Hertel</contributor>
  <contributor role="chair">Joseph W. Roggenbuck</contributor>
  <contributor role="chair">R. Bruce Hull</contributor>
  <date>1997-03-31</date>
  <type>Electronic Thesis or Dissertation</type>
  <format>application/pdf</format>
  <identifier> http://scholar.lib.vt.edu/theses/available/etd-3345131939761081/</identifier>
  <language>en</language>
  <rights>unrestricted</rights>
  <rights>I hereby grant to Virginia Tech or its agents the right to archive and to make available my thesis or dissertation
  in whole or in part in the University Libraries in all forms of media, now or hereafter known. I retain all proprietary rights, such
  as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or
  dissertation.</rights>
  <degree>
    <name>PHD</name>
    <level>doctoral</level>
    <discipline>Forestry</discipline>
    <grantor>Virginia Polytechnic Institute and State University</grantor>
  </degree>
</thesis>
```

4.4.2.2 Vyhledávání

NDLTD zajišťuje přístup k diplomovým pracem a disertacím partnerských organizací. Každá spolupracující instituce buduje vlastní archiv. Z těchto archivů se sklízí záznamy pomocí OAI Harvesteru a zpřístupňují se prostřednictvím souborného katalogu NDLTD. Data ze souborného katalogu jsou zpřístupňována prostřednictvím systému Virtua od firmy VTLS, přes vyhledávací systém Scirus, nebo přes experimentální vyhledávací systémy vyvíjené na Virginia Tech. Záznamy disertačních prací jsou sklízeny také do souborného katalogu OCLC, který dovoluje jejich vyhledávání přes několik rozhraní. Ze záznamů, které jsou sklizené do formátu Dublin Core, vedou odkazy na záznamy v souborném katalogu NDLTD.

Pro souborný katalog NDLTD je využíván integrovaný knihovní program Virtua společnosti VTLS Inc¹¹. Podporuje ukládání a sklizení metadatových záznamů přes protokol OAI ve formátech MARC, Dublin Core i ETD-MS. Zároveň umí pracovat s různými jazyky a to jak při ukládání tak při vyhledávání. Existují tedy lokální archivy vysokoškolských prací, ze kterých jsou záznamy sklizeny do centrální databáze (Chachra, 2006).

Obr. 4: Ukázka záznamu vyhledaného přes systém Virtua

full marc	
You searched NDLTD Collection - Subject: nutrition	
Author	Plum, Jane Meacham Jr.
Title	NUTRITION KNOWLEDGE ASSESSMENT OF PRESCHOOL CHILDREN
Publication	VT 1997-11-26
Summary Note	A game with food and nutrition related pictures was developed to provide an opportunity for a classroom teacher to interview preschool children for assessment of nutrition knowledge concepts. Specifically, knowledge of vegetable concepts which included identification of the food, the food group, the source, preparation methods and use by the body was measured. The assessment was administered to five groups of children (ages two and one-half to five years) in preschools and child care centers in Reston, VA by high school early childhood education students. The assessment was designed to meet current criteria that assessment be teacher administered, provide useful information to the teacher, reflect the typical activities of children in the classroom and be one of a variety of assessments used. The assessment was successfully administered by the high school students working as teacher aides. The assessment gave teachers useful information about each class and individual children which could be used in curriculum planning. The assessment fit into the usual classroom activities, in this case, a games and manipulatives learning center. The assessment provided more in-depth information about children's knowledge than multiple choice tests used in previous research on nutrition knowledge, because the teachers recorded children's responses to open-ended questions. Children's responses indicated elementary understanding of food and nutrition concepts suggesting that classroom teachers need to make greater use of the variety of nutrition education materials available.
Dissertation note	Master of Science (masters) -- VT.
Contributor	Brochetti, Denise
Contributor	Stewart, Daisy L.
Contributor	Hertzler, Ann A.
Institution	VT Human Nutrition, Foods, and Exercise
Subject	nutrition education
Subject	preschool
Subject	Assessment
Multimedia	Click to link to URL application/pdf http://scholar.lib.vt.edu/theses/available/etd-101397-204434/

¹¹ VTLS (Visionary Technology in Library Solutions) <http://www.vtls.com/>

Přes systém Virtua je možné vyhledávat pomocí prohlížení rejstříků, vyhledávání klíčových slov nebo přes expertní vyhledávání. Je možné zvolit si jazyk pro vyhledávací rozhraní. V současnosti se nabízí jedenáct jazyků. U konkrétního vyhledaného záznamu je možné zobrazit metadatový záznam nebo záznam ve formátu MARC. Záznam vyhledaný přes systém Virtua je na obrázku 4.

Další možností je použít vyhledávacího robota Scirus¹², který se zaměřuje na webové zdroje s vědeckým obsahem. Přes Scirus je v současnosti zpřístupňováno více než 237 500 plných textů diplomových a disertačních prací (Elsevier, 2007).

Přes seznam vyhledaných výsledků je možné zobrazit si konkrétní vybraný záznam přímo na stránkách univerzity, na které byla práce obhájena. Každá univerzita zobrazuje záznam jiným způsobem a nabízí jiné možnosti zobrazení nebo stažení metadatového záznamu. Přístup k plným textům závisí na každé univerzitě. Někdy je omezen na studenty a vyučující dané univerzity, jindy je např. před přístupem k plnému textu nutné souhlasit s nekomerčním využitím práce.

NDLTD klade důraz na vícejazyčné vyhledávání. Vícejazyčné vyhledávání řeší software, který používá pro překládání dotazů vícejazyčné slovníky. Dotaz v dalším jazyce je sice formulován velmi jednoduchým jazykem, ale napomůže vyhledávání v primárním jazyce databáze. Tyto překlady jsou používány především pro databáze v japonštině, korejštině nebo čínštině. Bohužel občas dochází i k chybným výsledkům vyhledávání, které jsou způsobené nesprávným překladem.

4.4.2.3 *Shrnutí informací k systému NDLTD*

Systém významnou měrou zasahuje do akademického prostředí v mnoha zemích různých kontinentů. Řeší také složité otázky autorských práv a vícejazyčného vyhledávání. Prosazuje vývoj informačních technologií a postupů na akademické půdě, zaměřuje se na studenty a způsoby jejich práce a tím ovlivňuje i vývoj praxe v oblastech, do kterých absolventi nastupují. NDLTD také doporučuje, aby samotní studenti připravovali metadata pro své práce. Podporuje tím vytváření a využití metadat, pochopení jejich důležitosti pro vyhledávání. Systém slouží jako podpora a vyvíjí software a postupy pro další projekty a systémy, které s ním spolupracují a vychází z jeho výsledků.

¹² Úvodní stránka vyhledávacího rozhraní Scirus je <http://www.ndltd.org/serviceproviders/scirus>.

NDLTD je významnou aktivitou v archivaci a zpřístupňování elektronických vysokoškolských prací. Celý systém má kvalitní technické a softwarové zázemí, které se neustále vyvíjí a zlepšuje.

Mezi největší problémy NDLTD patří zřejmě především to, že ze stran univerzit a vzdělávacích institucí není o spolupráci takový zájem, jaký byl očekáván. Díky tomu, že NDLTD onechává relativně velkou volnost jednotlivým institucím při tvorbě metadatových záznamů a volbě jejich formátu, jsou mezi záznamy patrné velké rozdíly.

4.5 WebArchiv



WebArchiv je jedním z českých projektů, které se nechaly inspirovat vývojem v oblasti metadat a to především metadatovým formátem Dublin Core. Původně byl WebArchiv dvouletým pilotním projektem financovaným z grantu Ministerstva kultury ČR. Probíhal v letech 2000-2001. Nositelem projektu byla Národní knihovna v Praze a řešitelkou Mgr. Ludmila Celbová. Na externí spolupráci se podíleli pracovníci Ústavu výpočetní techniky Masarykovy univerzity v Brně a Ikarosu. Oficiální název projektu byl *Registrace, ochrana a zpřístupnění domácích elektronických zdrojů v síti Internet*. WebArchiv je zkrácený název projektu, který se pro celý projekt vžil a provádí projekt i nadále (Vojtášek, 2001).

Projekt byl zaměřen na problematiku registrace, archivace a zpřístupnění elektronických zdrojů, zejména dostupných v síti Internet v doméně .cz. Elektronické zdroje tento projekt pojímá jako přirozenou součást národní publikační produkce a snaží se o jejich evidenci v rámci České národní bibliografie. Projekt si kladl za cíl vyřešit problematiku uchovávání elektronických zdrojů tak, aby byly v rámci národní bibliografie trvale dostupné, tzn. vytvořit digitální archiv bohemik.

V prvních fázích projektu byly především analyzovány zkušenosti zahraničních organizací a projektů se zpracováním elektronických zdrojů (Celbová, 2001). Dlouhodobá spolupráce byla navázána s Univerzitní knihovnou v Helsinkách (Vojtášek, 2002). V rámci projektu byly řešeny problémy nejen knihovnické, ale také legislativní a technické. Z knihovnického hlediska bylo nutné aplikovat na elektronické zdroje kritéria pro výběr zdrojů k registraci v národní bibliografii a zvolit příslušný formát, ve kterém budou zpracovávány záznamy o vybraných zdrojích (Celbová, 2001).

Projekt se od počátků potýkal s neuspokojivým stavem české legislativy v oblasti autorského zákona a zákona o povinném výtisku, takže spolupráci s vydavateli bylo nutné řešit formou individuálních dohod mezi Národní knihovnou a vydavateli, kteří byli ochotni spolupracovat (Vojtášek, 2002). Dle žádného zákona platného v ČR nemá Národní knihovna právo na povinný výtisk elektronického online zdroje. V Zákoně o neperiodických publikacích ani v Tiskovém zákoně není řešen povinný výtisk elektronického zdroje. Novela autorského zákona z roku 2006

dovoluje alespoň zpřístupnění díla pro veřejnost v prostorách poskytovatele, nedojde-li ke kopírování díla (Česko, 2006).

Z technického hlediska bylo nutné vyřešit tvorbu softwarových nástrojů pro vytváření metadat, autentifikaci zdrojů, stahování, archivaci a vyhledávání zdrojů a především zajistit jejich trvalou dostupnost s ohledem na stálý vývoj digitálních technologií (Celbová, 2001).

Výsledkem projektu byly podklady, které měly pomoci řešit právní zabezpečení získávání, archivace a zpřístupňování elektronických zdrojů publikovaných v síti Internet v doméně .cz a doménách .org a .com registrovaných českými subjekty (Celbová, 2002). Byly zpřístupněny některé softwarové nástroje zajišťující tyto procesy.

Šlo především o zpřístupnění generátoru metadat Dublin Core, který byl původně vyvinut v rámci Severského metadatového projektu (Nordic Metadata Project), a generátoru pro tvorbu jednotného identifikátoru URN (Uniform Resource Name) (Vojtášek, 2001). Další pomůckou zpřístupněnou na stránkách WebArchivu byl kalkulátor MD5. Slouží pro výpočet kontrolního součtu zadaného textového řetězce, který je v podstatě jednoznačným identifikačním údajem. Může vypočítat kontrolní součet pro celý zdroj (Žabička, 2002a).

Významným aktem bylo zahájení automatizovaného stahování a ukládání zdrojů z webu (Celbová, 2002). Tento postup se obecně označuje jako „sklizení dat“ (tzv. harvesting). Jde o automatické vyhledávání elektronických zdrojů na základě zadaných kritérií, jejich stahování do digitálního. Při zpřístupnění je velmi důležité datum vytvoření a stažení zdroje. Je možné prohlížet si webový prostor tak, jak vypadal v určitém období. Zdroje jsou stahovány na základě předem zadaných URL zdrojů, které jsou prohledávány dále a z jejichž odkazů se stroj dostane na další zdroje. Tímto způsobem je postupně prohledávána předem zadaná oblast Internetu, omezená např. doménou. Myšlenka sklizení dat vznikla ve Švédsku v roce 1996, později byla použita v rámci projektu NEDLIB (Networked European Depository Library)¹³ a projektu Severský webový archiv (Nordic Web Archive, NWA).

¹³ NEDLIB byl projekt z let 1998-2000 financovaný Evropskou komisí (v rámci programu Telematics for Libraries). Jeho hlavním cílem byl výzkum v oblasti systémů pro archivaci a zpřístupnění elektronických zdrojů. Do řešení tohoto projektu bylo zapojeno několik významných knihoven západoevropských zemí (Nizozemí, Francie, Německo, Norsko, Finsko, Itálie, Portugalsko a Švýcarsko) a vydavatelství Elsevier Science, Kluwer Academic a Springer-Verlag (NEDLIB, 1998).

WebArchiv přejal a upravil NEDLIB Harvester (Žabička, 2000). Jeho vývoj byl ovšem zastaven a tak byl v polovině roku 2004 nahrazen novým harvesterem Heritrix (Hutař, 2006).

Harvester slouží k procházení a sklizení volně dostupných bohemik dle teritoriálního hlediska, tzn. z domény .cz. Na odlišení kvalitních zdrojů ke sklizení byla vypracována metodika výběru zdrojů. Přednostně jsou na archivaci určené zdroje, u nichž se předpokládá odborná či umělecká hodnota. Z výběru jsou naopak vyřazeny zdroje reklamního charakteru. Sklízí se také některé zdroje dostupné prostřednictvím FTP. Není dořešena otázka dynamických webových stránek.

Po samotném sběru zdrojů následuje jejich aktualizace a doplňování. Postup probíhá stejným způsobem, ale stahují se pouze ty zdroje, které byly upraveny a změněny. Zda byl zdroj změněn, pozná harvester pomocí kontrolního součtu MD5.

Harvester ukládá do databáze zdroj a technická metadata (např. URL, informace z hlavičky HTTP). Pokud zdroj obsahuje popisný metadatový záznam, je vyextrahován a uložen také. Všechna metadata jsou převáděna a ukládána ve stejné znakové sadě (Žabička, 2000).

Úvodní projekt WebArchiv zajistil pouze stahování dat a budování jejich archivu. Neřešil jejich zpřístupnění (Žabička, 2002b). Následné fungování projektu WebArchiv bylo a je možné pouze díky dalším financím získaným z grantů. I v současnosti řeší projekt především problémy spojené se zajištěním trvalého přístupu k bohemikálním elektronickým zdrojům sítě Internet. Znamená to technické i legislativní zabezpečení získávání, zpracování, archivaci a ochranu těchto zdrojů, umožnění přístupu k nim při dodržení autorských práv. K tomu je nutné vytvářet a aktualizovat standardy a postupy a pokusit se vytvořit fungující spolupracující komunitu knihoven, informačních pracovišť i vydavatelů elektronických zdrojů.

Projekt WebArchiv je založen především na automatickém sklizení domény .cz. Sklizení probíhá ve dvou úrovních. Při plošné archivaci se sklízí prakticky celá zadaná doména omezená pouze určitými technickými parametry. Probíhá převážně automatizovně. Naproti tomu dochází při výběrové archivaci ke sklizení dokumentů vybraných podle určitých kritérií. V případě výběrové archivace je nezbytný určitý podíl intelektuální práce. Všechna kritéria nejsou ještě přesně ustálena. Mezi základní patří především kvalita obsahu zdroje, jeho bohemikální zaměření, příslušnost k doméně .cz. Dále může být zjišťováno, zda byl zdroj publikován

původně v digitální nebo tištěné formě, případně o jaký typ materiálu jde (monografie, seriál,...) (Hutař, 2006).

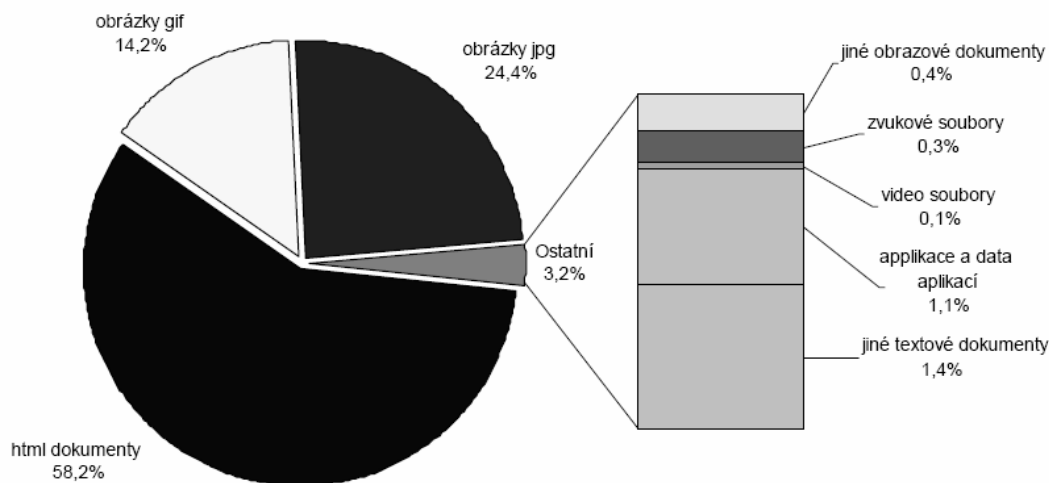
První sklízň dat byly provázeny problémy způsobenými mimo jiné nedostatečným technickým vybavením. Pro sklízění sloužil až do poloviny roku 2004 NEDLIB Harvester. Byl nahrazen novým harvesterem Heritrix. (Hutař, 2006). Zdroje, na které má Národní knihovna smlouvy o zpřístupnění s vydavateli, jsou sklízeny několikrát ročně. Ostatní zdroje jsou sklízeny v rámci celoplošných sklízň domény .cz. Problémy jsou ale se sklízěním domén druhé úrovně a podchyceny nejsou ani bohemika mimo doménu .cz. Tabulka 6 ukazuje objem sklizených dat v jednotlivých letech (Celbová, 2006).

Tab. 6: Přehled sklízň domény .cz (Celbová, 2006)

Rok	Počet sklizených souborů	Nekomprimovaná velikost [GB]	Délka sklízění [dny]	Počet domén druhé úrovně	% z tehdy registrovaných domén druhé úrovně
2001	3.015.057	104	21	41.322	38%
2002	10.249.302	307	93	79.022	69%
2004	32.141.575	1.034	204	101.378	75%
2005	9.336.123	247	12	4.795	2%
2006	72.378.019	3.416	40	196.880	74%

Graf na obrázku 5 znázorňuje zastoupení jednotlivých jednotlivých typů souborů v archivu WebArchiv. Jak je vidět, většinu dokumentů tvoří HTML dokumenty, zastoupeny jsou v menší míře i jiné formáty (Celbová, 2004).

Obr. 5: Graf zastoupení typů souborů v archivu WebArchiv (Celbová, 2004)



4.5.1 Popisné jednotky

Popisnými jednotkami jsou monografie, seriály i jednotlivé články. Granularita není jednotná. Záleží na dohodách s vydavateli, zda jsou popisovány jednotlivé články nebo celé časopisy (Vojtášek, 2002).

Tvůrcům záznamů je k dispozici webový generátor. Ten je možné použít nejen k vytváření metadat pro zdroj, ale následně také pro natažení metadat, která byla již vytvořena, do formuláře. Údaje ve formuláři je možné zobrazovat jako náhled, ve formátu XHTML a RDF/XML. Tato zobrazení lze kombinovat ještě s popisem v nekvalifikovaném a kvalifikovaném Dublin Core.

4.5.2 Metadatové záznamy

Pracovníci WebArchivu vytvářejí ručně dva typy záznamů. Prvním jsou záznamy v systému Aleph ve formátu MARC 21 a druhým metadatové záznamy ve formátu Dublin Core.

Pro zdroje, vybrané na základě stanovených kritérií, vytváří záznamy v programu Aleph ve formátu MARC 21. Záznamy byly do roku 2006 uloženy v samostatné bázi, v současnosti se ale stávají součástí centrální báze Národní knihovny (Jelínková, 2006). Tyto záznamy slouží jako součást kulturního dědictví a jsou registrovány zároveň v České národní bibliografii (Celbová, 2004). Přes pole 856 je zajištěno propojení těchto záznamů jak přímo s aktuální webovou stránkou tak se záznamem v digitálním archivu (Jelínková, 2005).

Pro vytváření metadatových záznamů se používá formát Dublin Core. Záznamy jsou vytvářeny ručně a to pouze pro zdroje, na jejichž zpřístupnění má WebArchiv smlouvu s dodavateli. V dohodách, které Národní knihovna a WebArchiv uzavírají s vydavateli a autory online zdrojů, je uvedeno, že vydavatelé se zavazují vkládat do webových stránek do hlavičky metadatové záznamy, které pro ně pracovníci WebArchivu vytvoří. Vydavatelé mají zajištěnu vyhledatelnost daných záznamů v archivu WebArchiv. Vložením metadat do zdroje je zdroj lépe vyhledatelný přímo v síti Internet (Jelínková, 2005).

Metadatový záznam ve formátu Dublin Core si může pomocí generátoru, který je přístupný na stránkách WebArchivu, vygenerovat kdokoli. Následně je možné vložit záznam do webové stránky a využít pro vyhledávání (Celbová, 2004).

Generátor obsahuje řadu pomůcek. Kromě samotného formuláře, do kterého se vyplňují údaje, je doplněn rolovacími nabídkami pro výběr kvalifikátorů, typů a formátů zdroje, použitých řízených slovníků a heslářů a třídění a jiných údajů. (Generátor metadat, 2001).

WebArchiv používá prvky Dublin Core. V rámci projektu byly prvky přeloženy do češtiny. Ve formuláři, který se používá pro generování metadatových záznamů jsou uvedeny české výrazy. Výsledné záznamy ovšem obsahují prvky v originálu. Prvky i s kvalifikátory jsou uvedeny v příloze 4 (Generátor metadat, 2001). Jde o standardní prvky Dublin Core. Jedinou odlišností jsou používané české tezaury a klasifikace.

4.5.3 Vyhledávání

Koncem roku 2005 zpřístupnil WebArchiv volně dostupnou část archivu přes jednoduché rozhraní pro vyhledávání na svých stránkách. Zpřístupněny jsou zdroje vydavatelů, kteří uzavřeli s Národní knihovnou smlouvu o poskytování elektronických online zdrojů. Vyhledávat je možné pomocí zadaného slova, víceslovného výrazu, přesné fráze nebo URL. Vyhledané výsledky je možné omezit rozpětím let.

Obr. 6: Ukázka metadat k vyhledané webové stránce uložené ve WebArchivu

The screenshot shows the WebArchiv interface. At the top, there is a search bar with the URL 'http://www.nkp.cz/tisk/dalimil_kronika/index.htm' and a search button. Below the search bar, there is a timeline slider showing the years 2001 and 2011. The main content area displays the metadata for the selected document. The metadata is organized into sections: Archive Identifier (aid), Url, Time of archival, Last modified time, Mime-type, File Status, Content Checksum, and HTTP Header. The HTTP Header section shows the date, time, server, and other technical details.

Uri: Go Search Go

Viewing version 3 of 3
4. Únor 2006, 15:24

2001 2011

Resolution: Years Auto: ☐ Metadata: ☒ [Nápověda](#)

Metadata

Archive Identifier (aid)
11954362/IAH-20060214152316-00000-harvester

Url
http://www.nkp.cz/tisk/dalimil_kronika/index.htm

Time of archival
20060214152428

Last modified time
20050413104414

4253

Mime-type
text/html

File Status
online

Content Checksum
X5WWY4LUXVN37X5M2NVNIREZPN4EPKQ6

HTTP Header
Date: Tue, 14 Feb 2006 15:24:27 GMT Server: Apache Last-Modified: Wed, 13 Apr 2005 10:44:14 GMT ETag: "24ee1c-109d-425cf7fe" Accept-Ranges: bytes Content-Length: 4253 Connection: close Content-Type: text/html

Přes odkaz „More from this site“, který je dostupný u každého vyhledaného záznamu, probíhá opakované spuštění dotazu s omezením na doménu vybraného zdroje (např. <http://www.nkp.cz>). U každého záznamu je možné zobrazit přehled jeho dostupných verzí, které byly sklizené. Je možné zobrazit jak vybranou verzi zdroje s informací o časových verzích, tak metadata k této stránce. Bohužel, jak ukazuje obrázek 6, jde o metadata technická, která se vztahují k technickému procesu sklizení pomocí harvesteru.

Pro indexování zdrojů se používá fulltextový nástroj pro indexaci webu NutchWAX (Nutch - Web Archive Extension), který je rozšířený o podporu vyhledávání v časových verzích. Index, který NutchWAX vytváří, používá uživatelské rozhraní WERA (Web Archive Access), se kterým pracují uživatelé (Žabička, 2005).

4.5.4 Shrnutí informací k projektu WebArchiv

Projekt WebArchiv se snaží o zachování rozsáhlé části národního kulturního dědictví, která je dostupná prostřednictvím sítě Internet. Primárně se zaměřuje na původní elektronické online zdroje, jejichž životnost je často velmi krátká. Cílem projektu je archivovat a zpřístupnit tyto zdroje současným i budoucím generacím. Neopomíjí ani skutečnost, že řada elektronických online zdrojů je pochybné kvality. Snaží se proto soustředit na zdroje, u kterých se předpokládá kulturní, společenský nebo odborný přínos a zaměření na národní otázky. Protože však nelze předem tušit, jaké informace budou mít hodnotu v budoucnosti, probíhá kromě výběrového sklizení dat také sklizení celé národní domény.

Od svých počátků se potýká s řadou finančních i legislativních problémů a omezení. Přesto je patrný jeho vývoj a posun v řešení problémů. Významný je i fakt, že WebArchivu přispívá částí ručně vytvořených záznamů do České národní bibliografie a předpokládá se, že se jeho archiv stane v roce 2007 součástí připravované Digitální knihovny (Hutař, 2006).

Vzhledem k tomu, že WebArchiv používá pouze prvky formátu Dublin Core, neměl by nastat problém při konverzi, výměně nebo sdílení záznamů nejen s dalšími českými institucemi, ale ani při spolupráci se zahraničními organizacemi. Prvky Dublin Core jsou používány standardně. Problémy by mohly nastat pouze v případě polí, jejichž obsah je ošetřen pomocí řízených slovníků, které se používají pouze

v České republice (Český pedagogický tezaurus, Český teologický tezaurus apod.). WebArchiv připravuje také implementaci standardů METS a OAI-PMH (Matějka, 2006).

O úspěšnosti projektu WebArchiv svědčí i ocenění, které získal v roce 2006 na konferenci Inforum. O udělení této ceny hlasují účastníci konference. Vzhledem k tomu, že jde o odbornou konferenci s mezinárodní účastí, které se vždy účastní řada významných osobností z oboru informačních služeb, technologií a knihovnictví, lze usuzovat, že si odborná veřejnost zásluhy tohoto projektu uvědomuje a oceňuje je.

Více o projektu WebArchiv je možné najít na stránkách věnovaných projektu a také v diplomových pracích Markéty Škodové (Škodová, 2005), Lenky Jelínkové (Jelínková, 2006) a Lukáše Matějky (Matějka, 2006).

4.6 Projekt ANL FULL

Databáze ANL FULL, která obsahuje plné texty vybraných článků z novin a časopisů Anopress je také výsledkem několika navazujících projektů. V roce 1999-2003 probíhal výzkumný záměr *Propojení analytických záznamů s plnými texty a optimalizace zpřístupnění plných textů* a v letech 2000-2004 programový projekt *Souborná databáze kooperačního systému článkové bibliografie – optimalizace integrace a správy heterogenních dat*. Oba projekty se zabývají otázkou získávání, zpracování a zpřístupnění českých plných textů a řeší aplikaci metadatového formátu Dublin Core na popis elektronických článků z českých novin a časopisů s ohledem (Anděrová, 2002c).

Báze ANL FULL je úzce propojená s další databází ANL. Obě báze jsou výběrové a obsahují převážně odborné tituly. Báze ANL pokrývá seriály vydávané Akademií věd ČR, vysokými školami i dalšími institucemi v České republice, dále noviny a kulturně-politické a populárně naučné časopisy. Báze ANL FULL zachycuje pouze noviny a kulturně-politické časopisy a navíc časopis Národní knihovna. Rutinně se zpracovávají články ze 17 titulů, výběrově z cca 37 titulů. Do obouází jsou zařazovány pouze odborné články a články, které nesou nějakou dokumentární, uměleckou, politicko-společenskou nebo jinou hodnotu. Tematicky jsou zastoupeny obory humanitní a společenské, ekonomika, obchod a finance, matematika a přírodní vědy, lékařství, technika a volný čas (Anděrová, 2002b).

Báze ANL FULL byla plněna fulltextovými články z databáze TamTam provozované firmou Anopress. Pro zpracování a zpřístupnění plných textů byla vyvinuta poloautomatická linka extrakce/indexace, která využívá TTDE (TamTam Data Extractor). V rámci této linky bylo zajištěno částečné stažení některých údajů a dále bylo nutné dopracovat záznamy především z věcného hlediska a propojit je na autority Národní knihovny (Anděrová, 2005b).

Zpracování v Národní knihovně pracovníky ANL FULL probíhalo v on-line režimu. Katalogizátor si po přihlášení do systému stáhl článek na server Národní knihovny. Po dekompresi pracovník načel pomocí technologie TTDE do formuláře údaje k příslušnému článku z hlavičky daného článku. Spolu s načtenými údaji z databáze TamTam se vygenerovaly některé údaje, které jsou nezbytné především

pro záznam ve formátu MARC. Automaticky se vygeneroval také souhrn článku (první věta textu), dynamické URL, URN, SICI. Generované údaje byly v souladu s pravidly AACR2. Pracovník doplnil ve formuláři údaje věcného popisu (MDT, předmětové kategorie, předmětová hesla, klíčová slova), typ zdroje, případně zkontroloval ostatní údaje a záznam odeslal. Záznam bylo možné editovat a kontrolovat v několika záložkách formuláře. Kromě základní záložky s jmennými a věcnými údaji bylo možné zobrazit si pouze jmenné údaje, záznam MARC, Dublin Core, zobrazovací hlavičku a indexovací hlavičku. Zobrazovací hlavička sloužila k zobrazení údajů v hlavičce plného textu a indexovací hlavička pro indexaci uložených údajů v databázi ANL FULL (Anděrová, 2002a).

Z údajů ve formuláři se automaticky vytvořilo několik záznamů. První byl záznam hlavičky. Dále byl automaticky vytvořen záznam ve formátu MARC. Třetím záznamem, který byl na základě údajů ve formuláři vytvořen, byl metadatový záznam Dublin Core ve formátu HTML. Vytvářel se také kvalifikovaný a nekvalifikovaný záznam Dublin Core v XHTML a v XML/RDF. Záznam ve formátu MARC byl odeslán do bibliografické báze ANL do systému Aleph. Zobrazovací hlavička se ukládala do plného textu a spolu s ním byla uložena v databázi ANL FULL. Údaje z indexovací hlavičky byly zaneseny do indexů ANL FULL pro vyhledávání v systému Topic. Vygenerovaný záznam Dublin Core byl uložen do plného textu a spolu s ním do zvláštního adresáře (Anděrová, 2002a).

Tento proces měl zajistit propojení záznamů v obou bázích a zároveň stálou dostupnost plného textu. Plné texty byly navíc indexovány fulltextově, aby bylo umožněno nejen vyhledávání pomocí vytvořených záznamů ale také fulltextové hledání.

Báze ANL FULL využívala sofistikovaný software pro automatickou extrakci a vyhledávání firmy Verity nazvaný TOPIC¹⁴, který umožňuje tzv. plnotextové pojmové vyhledávání. Při něm se vytváří pojmově příbuzné „shluky“, které se následně využívají při vyhledávání. Umožněno je také hodnocení relevance, shlukování dokumentů podle společného kontextu, nebo vytváření automatických anotací. Definované topiky je možné chápat jako pojmové hierarchické větve, které vytváří znalostní databázi (Anděrová, 2005c).

¹⁴ Novější verze systému TOPIC jsou označovány jako Portal One.

V roce 2004 nebyl přijat projekt předkládaný v rámci 1N nazvaný *Kooperační systém článkové bibliografie - nástroje pro zpracování a zpřístupnění odborných informačních zdrojů, porovnání výsledků výzkumu a vývoje v ČR se světem*, který měl zajistit vytvoření nového konceptu Kooperačního systému článkové bibliografie, úpravy struktur používaných třídění, obnovení technického a softwarového zázemí projektu.

Ve stejném roce došlo k nahrazení zpracování v lince TTDE zapisováním záznamů přímo do šablon v systému Aleph. V roce 2005 následovalo zastavení doplňování plných textů do databáze ANL FULL, takže odkazy za záznamů vedou pouze na plné texty v databázi Anopress. V současnosti obsahuje databáze ANL FULL cca 130 000 plných textů (Anděrová, 2005a).

Druhá databáze ANL nebyla plněna pouze záznamy zpracovávanými do báze ANL FULL, ale vzniká v rámci Kooperačního systému článkové bibliografie (tzv. KOSABI), do kterého jsou kromě Národní knihovny zapojeny krajské knihovny, Moravská zemská knihovna a některé odborné knihovny České republiky. V rámci tohoto systému je vytvářena souborná databáze ANL¹⁵, která tvoří jednu z řad České národní bibliografie nazvanou *Články v českých novinách, časopisech a sbornících* (Anděrová, 2000). Přispívající knihovny musí data často konvertovat, aby bylo možné je do databáze ANL zařadit. V roce 2004 byla celá databáze konvertována z formátu UNIMARC do formátu MARC 21. Ve stejném roce se začal používat pro popis také konspekt (Anděrová, 2005c).

Báze ANL FULL i ANL jsou zahrnuty jako zdroje v Jednotné informační bráně. Bázi ANL je možné prohledávat. Báze ANL FULL má jistá omezení. Vyhledávání je možné z interní sítě Národní knihovny a přístupné jsou i plné texty článků (Anděrová, 2002b).

Je-li v bázi ANL prostřednictvím Jednotné informační brány vyhledán záznam článku, který je zároveň obsažen v bázi ANL FULL, uživatel může přes odkaz do této plnotextové databáze přejít.

¹⁵ Databáze ANL je v systému Aleph. Záznamy přispívajících knihoven, které nejsou zpracovávány v Národní knihovně České republiky jsou konvertovány do systému Aleph. Databáze obsahuje v současnosti přibližně 740 000 bibliografických záznamů. Zpracovávány jsou články z 537 titulů seriálů s retrospektivou od roku 1990 (Anděrová, 2002c).

4.6.1 Popisné jednotky

V rámci ANL FULL je jednoznačně stanovena popisná jednotka. Jsou jí vždy jednotlivé články. Záznamy se vztahují k plným textům článků, recenzí, komentářů a sloupků z českých novin a časopisů z databáze TamTam. Články byly vybírány podle svého obsahu. Přednostně byly popisovány články s odborným, společenskovedním a kulturním obsahem.

4.6.2 Metadatové záznamy

Báze ANL FULL používá 14 prvků formátu Dublin Core. Prvek *dc.relation* (vztah) nebyl využit. Naopak byl přidán prvek pro podnázev *dc.subtitle* a bylo nadefinováno dalších devatenáct prvků. Tyto prvky podrobněji člení údaje věcného popisu, čehož následně využívá k přesnějšímu vyhledávání systém Topic (Anděrová, 2005b). Vznikl tzv. ANL Core.

Od standardních prvků Dublin Core, které jsou uvozeny prefixem *dc*, jsou prvky ANL Core odlišeny prefixem *anl* (Národní knihovna, 2000). Prvky ANL Core jsou vyjádřeny buď v rámci prvku Dublin Core za tečkou, nebo jsou uvedeny jako opakování prvku Dublin Core. Přehled prvků, které ANL FULL používá, je uveden v tabulce 7 (Anděrová, 2004).

Jak z tabulky vyplývá, některé prvky Dublin Core jsou použity vícekrát pro různé údaje. Označení *<dc.title>* slouží pro název článku i pro název části nebo článku na pokračování. Pro jemnější specifikaci názvu části nebo článku na pokračování lze užít také prvek ANL Core *<anl.tipart>*. Obdobně je tomu u prvku *<dc.source>*, který lze užít pro určení názvu celého zdroje i pro číslo ISSN. Pro číslo ISSN lze použít také prvek *<anl.issn>*. Pro věcný popis bylo definováno několik prvků ANL Core. Jde např. o *<anl.topic>*, *<anl.mdt>*, *<anl.person>*, *<anl.corporation>* nebo *<anl.act>*, které lze zařadit pod prvek Dublin Core *<dc.subject>*. Nutnost užší specifikace je jedním z důsledků obecné specifikace prvků Dublin Core.

Tab. 7: Metadatové prvky použité v databázi ANL FULL

Význam prvku	Prvky Dublin Core	Prvky ANL Core	Schéma, kvalifikátor	Prefix v češtině	Prefix v angličtině
Název	dc.title			naz	ti
Název části/Název článku na pokračování	dc.title	anl.tipart		nan	tip
Podnázev	dc.subtitle			pdn	sbt
Autor (primární odpovědnost)	dc.creator.personalName			auh	aup
Autor (primární odpovědnost) - korporace	dc.creator.corporateName			auh	aup
Přispěvatel (další autor/sekundární odpovědnost)	dc.contributor			auv	aus
Vydavatel	dc.publisher			vyd	pub
Zdroj (název, roč., číslo, datum, strany)	dc.source			zdr	src
Název části zdroje		anl.so-part		zdn	sop
Rubrika		anl.cap		rbr	cap
Zdroj (číslo ISSN)	dc.source	anl.issn	ISSN	issn	issn
Svazek/ročník		anl.vol		roc	vol
Číslo		anl.nr		cis	nr
Datum vydání	dc.date.issued		W3C-DTF	dat	dat
Datum zpracování	dc.date.modified		W3C-DTF	dac	dac
Typ/druh zdroje	dc.type		DCMI	typ	typ
Formát zdroje	dc.format				
Jazyk	dc.language		RFC1766		
Práva	dc.rights				
Identifikátor	dc.identifier		SICI, URN		
Identifikátor (identifikační číslo článku)		anl.id		id	id
Pokrytí prostorové	dc.coverage.spatial		GAC	zek	gac
Pokrytí časové	dc.coverage.temporal			cak	tpc
Předmět (klíčová slova)	dc.subject	anl.key		kls	key
Předmět (téma jako předmět)	dc.subject		PHNK	phnk	phnk
		anl.sub		sub	sub
Předmět (geografické jméno jako předmět)	dc.subject		PHNK	phnk	phnk
		anl.geo		geo	geo
Předmět (předmětové kategorie ANL)	dc.subject	anl.topic	ANL-kategorie	top	top
Předmět (osoba jako předmět)	dc.subject		PHNK	phnk	phnk
		anl.person		oso	per
Předmět (korporace jako předmět)	dc.subject		PHNK	phnk	phnk
		anl.corporation		kor	cor
Předmět (akce jako předmět)	dc.subject		PHNK	phnk	phnk
		anl.act		akc	act
Předmět (název jako předmět)	dc.subject		PHNK	phnk	phnk
Předmět (dílo jako předmět)		anl.wrc		dil	wrc
Předmět (autor/dílo jako předmět)	dc.subject		PHNK	phnk	phnk
		anl.adl		adi	awr
Předmět (notace MDT)	dc.subject	anl.mdt	MDT	mdt	udc
Předmět (notace DDT)	dc.subject	anl.ddc	DDT	ddc	ddc
Předmět (Konspekt)	dc.subject	anl.kns	Konspekt	kns	cns
nepoužívá se	dc.relation				

V databázi ANL FULL bylo využito sedmnáct řízených slovníků. Jde o řízené slovníky autorů, názvů zdrojů, jejich ročníků a čísel, dat vydání, předmětových kategorií, předmětů, míst, věcných a geografických hesel, osob, korporací, akcí, klíčových slov, zdrojů/děl, typů článků, MDT, geografických a časových vymezení (Národní knihovna, 2000). Rejstříky slouží k vyhledávání. V průběhu řešení projektu byla řešena aktualizace rejstříků a deduplikace obsažených hesel. Zavedlo se také používání databáze Národních autorit. Rejstřík

předmětových kategorií byl napojen na notace MDT, které se při výběru předmětové kategorie doplňovaly automaticky.

4.6.3 Vyhledávání

Na webových stránkách ANL FULL pod sekci Vyhledávání jsou dostupné rejstříky a tři rozhraní pro vyhledávání (základní formulář, rozšířený formulář a rozšířený formulář s tématy). K dispozici jsou rejstříky: Autoři; Název zdroje; Zdroj, ročník, číslo; Datum vydání; Předmět, kategorie; Předmět; Místo; Hesla; Osoby; Korporace; Akce; Klíčová slova; Dokument/dílo; Typ článku; MDT; Geograf. vymezení; Chronol. vymezení (Národní knihovna, 2000).

Výsledky vyhledávání se zobrazují v okně rozděleném na dvě části. Levá část obsahuje seznam vyhledaných záznamů, v pravé části se zobrazuje náhled vybraného záznamu. Je-li uživatel v interní síti Národní knihovny, má přístup k plnému textu.

Příklad 10: Metadatový záznam vyhledaný v bázi ANL FULL v kvalifikovaném formátu DC

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" standalone="yes"/>
<meta name="DC.Title" content="Miliardy na družstevní byty"/>
<meta name="DC.Creator.personalName" content="Pravec, Josef"/>
<meta name="DC.Creator.personalName" content="Skalková, Olga"/>
<meta name="DC.Subject" scheme="ANL-KATEGORIE" content="veřejná správa"/>
<meta name="DC.Subject" scheme="ANL-KATEGORIE" content="ekonomika"/>
<meta name="DC.Subject" scheme="MDT" content="351.778.5"/>
<meta name="DC.Subject" scheme="MDT" content="334.736"/>
<meta name="DC.Subject" scheme="MDT" content="(437.3)"/>
<meta name="DC.Subject" scheme="PHNK" content="bytová politika-Česko-r. 2004"/>
<meta name="DC.Coverage.spatial" content="Česko"/>
<meta name="DC.Coverage.temporal" content="r. 2004"/>
<meta name="DC.Subject" scheme="PHNK" content="bytová družstva-Česko-r. 2004"/>
<meta name="DC.Coverage.spatial" content="Česko"/>
<meta name="DC.Coverage.temporal" content="r. 2004"/>
<meta name="DC.Description.abstract" content="Na jeden byt by družstva od 1. července příštího roku mohla dostávat nenávratnou stotisícovou dotaci a zvýhodněný státní úvěr až do 700 tisíc korun za roční tříprocentní úrok"/>
<meta name="DC.Publisher" content="Economia"/>
<meta name="DC.Date.issued" scheme="W3C-DTF" content="2004-10-06"/>
<meta name="DC.Type" scheme="DCMI" content="text"/>
<meta name="DC.Type" content="article"/>
<meta name="DC.Type" content="článek"/>
<meta name="DC.Format" scheme="IMT" content="text/html"/>
<meta name="DC.Identifier" content="http://full.nkp.cz/nkdb/view/mkdoc.asp?p1=HN20041006000001"/>
<meta name="DC.Identifier" scheme="SICI" content="0862-9587(20041006)48:195<1:HN20041006000001>3.0.CO;2-K"/>
<meta name="DC.Identifier" scheme="URN" content="URN:NBN:cz-nkHN20041006000001"/>
<meta name="DC.Source" content="Hospodářské noviny roč. 48, č. 195 (2004-10-06), s. 1"/>
<meta name="DC.Source" scheme="ISSN" content="0862-9587"/>
<meta name="DC.Language" scheme="RFC1766" content="cs"/>
<meta name="DC.Coverage.spatial" scheme="GAC" content="e-xr---"/>
<meta name="DC.Rights" content="Economia"/>
<meta name="DC.Rights" content="Anopress"/>
<meta name="DC.Date.modified" scheme="W3C-DTF" content="2004-10-06"/>
```

K metadatovému záznamu je možné dostat se přes ikonu jablka u titulů v seznamu vyhledaných článků. Další ikony u vyhledaných záznamů nabízí zobrazení záznamu s odkazy, přes které je možné spustit vyhledávání v příslušných

rejstřících. Metadatový záznam je možné zobrazit v jednom z dostupných formátů. Nabízí se formát UNIMARC a následně několik variant metadatového záznamu ve formátu Dublin Core – kvalifikované i nekvalifikované ve formátech XHTML i XML.

Příklad 10 ukazuje metadatový záznam v kvalifikovaném formátu Dublin Core zapsaný v XHTML vyhledaný v bázi ANL FULL. Jak je vidět, záznam odpovídá pravidlům formátu Dublin Core. Některé prvky jsou pouze podrobněji rozpracovány. Používají se také české klasifikace a číselníky.

Jiný záznam ukazuje příloha 5. Jde o kvalifikovaný záznam ve formátu RDF, který popisuje recenzi. Záznam je strukturovaný. Údaje jsou vloženy do příslušných prvků. V záznamu je uplatněno uzavírání výčtu prvků do RDF kontejneru typu <bag>. V kvalifikátorech jsou určena schémata, podle kterých byl zapsán obsah prvků. Jednotlivé položky jsou uvozeny pomocí prvku .

4.6.4 Shrnutí informací k projektu ANL FULL

Protože je v databázi ANL FULL kladen kvůli následnému vyhledávání důraz na věcný popis a identifikaci zdroje (celého časopisu, jeho ročníku, čísla apod.), nedostačovala obecná specifikace prvků Dublin Core. Prvky Dublin Core byly použity a rozšířeny o další prvky, které především v oblasti věcného popisu podrobněji rozlišují významu prvků.

Zajímavá byla i linka zpracování TTDE, která byla částečně automatizovaná. Na základě údajů vyextrahovaných ze zdroje a dalších údajů zadaných katalogizátorem vznikalo více záznamů, které měly různé použití.

Současný stav, kdy se ANL FULL potýká s finančními problémy a plnění databáze ANL FULL je zastavené, není optimální. V roce 2006 byly ANL FULL poskytnuty pouze základní finanční prostředky na údržbu serveru a řada naplánovaných cílů nebyla realizována (Anděrová, 2005a). Budoucnost ukáže, zda bude projektu znovu dovoleno pokračovat.

5 HLAVIČKY TEI A JEJICH APLIKACE

5.1 Formát TEI



Formát Iniciativy pro kódování textů (Text Encoding Initiative, TEI, <http://www.tei-c.org/>) je jedním z nejstarších metadatových schémat. Práce na tomto formátu započaly v roce 1987 v rámci mezinárodního výzkumného projektu vedeného Asociací pro využití počítačů v humanitních vědách ACH (Association for Computers in Humanities), k níž se přidaly Asociace pro využití počítačů v literatuře a lingvistice ALLC (Association for Literary and Linguistic Computing) a Asociace pro počítačovou lingvistiku ACL (Association for Computational Linguistics) (MIT Libraries, 2004). V roce 2000 vzniklo mezinárodní neziskové konsorcium TEI, které nadále udržuje a rozvíjí formát TEI (Text Encoding Initiative, 2007b).

Výkonná složka sídlí ve Virginii v USA. Hostováno je TEI na univerzitách v Oxfordu, Bergenu a na Brownově univerzitě. Spoluúčastní se i další pracovní skupiny a instituce. Mezi řadu sponzorů tohoto projektu patří mimo jiné Mellonova nadace, Americký fond pro humanitní vědy nebo Evropská komise a její XIII. ředitelství (Text Encoding Initiative, 2007b).

Výsledky práce původního projektu a následně konsorcia na tvorbě formátu začaly být zveřejňovány v roce 1990, kdy byl k dispozici první koncept formátu TEI. Následoval další pod označením TEI P2. V roce 1994 byl publikován manuál pod označením TEI P3. Od roku 2002 byla verze TEI označovaná jako TEI P4. V roce 1995 vyšla zjednodušená verze TEI Lite označovaná jako TEI U5. V současnosti je aktuální TEI P5 verze 0.6 z března 2007. Vydání verze 1.0 se očekává během roku 2007. Původně vycházel formát TEI ze SGML. Nyní se již orientuje na standarty XML, W3C Scheme nebo Relax NG Scheme¹⁶. Schémata a definovaná DTD zajišťují, že v dokumentech budou použity pouze definované prvky a jejich vlastnosti. Nicméně definované DTD není závazné. Konsorcium TEI upozorňuje, že jde o návrh a odhad, jak by mělo takové DTD vypadat, aby vyhovovalo potřebám většiny lidí/institucí (Burnard, 2005).

¹⁶ Relax NG – Regular Expression Language for XML New Generation

TEI je rozsáhlý formát pro popis všech typů textových i netextových objektů. Používá se především v oblastech literatury, lingvistiky a humanitních věd. Je určen hlavně pro muzea, knihovny a vydavatele (Romary, 2004).

5.1.1 Popisné jednotky

Hlavička TEI se vztahuje vždy k jedné popisné jednotce. Záleží ovšem na tom, co je jako popisná jednotka vnímáno. Existuje-li zdroj, který je složen z mnoha menších zdrojů, může být hlavička TEI použita k popisu jak celého zdroje tak i jednotlivých menších vnořených zdrojů. Díky tomu se může stát, že jeden dokument bude obsahovat jednu souhrnnou hlavičku a uvnitř jednotlivé hlavičky např. ke kapitolám textu (Humanities Text Initiative, 1993).

5.1.2 Struktura formátu TEI

Formát TEI je velmi obsáhlý, obsahuje velké množství prvků, ale pouze několik jeho prvků je povinných. TEI se skládá z řady částí/modulů. Některé moduly jsou základní a je nutné je použít, jiné jsou doplňkové.

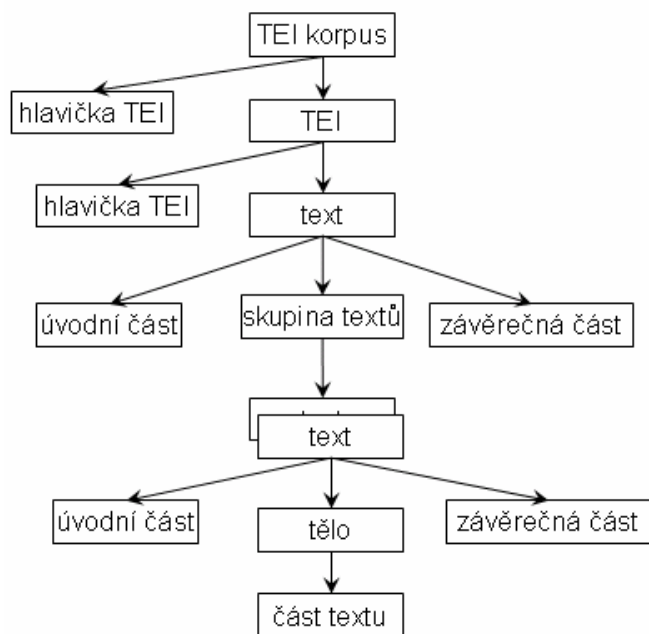
Základem je infrastruktura, která definuje všechny třídy (skupiny) prvků, které jsou dále použity. Tato část musí být vždy přítomna. Aby text vyhovoval všem pravidlům TEI, měl by obsahovat i druhou část, kterou tvoří hlavička TEI. Ta obsahuje všechny prvky a údaje sloužící bibliografickému popisu zdroje. Třetí částí je strukturovaný text zdroje. Nad těmito třemi základními kameny TEI mohou být další moduly a specializované aplikace – např. textová kritika, jazykové korpusy, provázání údajů apod. (Text Encoding Initiative Consortium, 2007a).

TEI se používá pro popis celého zdroje. Informace o zdroji nese hlavička TEI, ostatní prvky TEI jsou umístěny v textu. Prvky umístěné v textu mohou vypovídat o způsobu zobrazení určeného textu (strukturální prvky), ale mohou se vztahovat také k významu označeného textu (sémantické prvky). Každý prvek může mít nějaké vlastnosti (atributy). TEI může být tedy využito jak k zachycení struktury tak obsahu textu (Burnard, 2005).

TEI může popisovat jednotlivé zdroje nebo zdroj složený z více textů. V případě složeného textu se hlavní část opakuje pro každý jednotlivý text.

Schéma na obrázku 7 naznačuje základní strukturu popisu (Romary, 2004). Popis je složen z popisu textu a hlavičky, která nese informace o zdroji/textu. Text

Obr. 7: Základní struktura popisu TEI



Strukturu záznamu ve formátu TEI tvoří prvky a jejich vlastnosti sdružené do skupin. Prvek je textová jednotka (odstavec, název, autor apod.). Vlastnost prvku vypovídá o podrobnějším výskytu prvku a může být opět tvořena párem - vlastnost/hodnota. Prvky jsou rozděleny do čtyř skupin:

- jádro - prvky potřebné ve všech zdrojích. Jde např. o prvky zaznamenávající autory, korporace, názvy, předmětová třídění, odkazy v jednom zdroji i ze zdroje na jiný zdroj
- základní skupina - prvky vyhrazené pro podrobné členění typů zdrojů
- doplňková skupina - prvky určené pro specializaci a detailnější popis textu
- pomocná skupina - prvky se specifickým účelem

Každá organizace může vytvořit vlastní DTD TEI (Heery, 1996). Jedním z velmi rozšířených DTD založených na TEI je TEI Lite. Pravidla tvorby záznamu ve formátu TEI jsou velice flexibilní. Je možné kombinovat soubory prvků a jejich vlastností tak, aby byly uspokojeny potřeby popisu konkrétního zdroje (Burnard, 1998).

5.1.3 Hlavičky TEI

Hlavičky TEI jsou součástí rozsáhlého formátu TEI. Hlavičky obsahují detailní bibliografický popis zdroje ve formě metadat. Používají se k textovým analýzám či k bibliografické kontrole. Specifickou skupinu hlaviček TEI tvoří nezávislé hlavičky TEI, které se používají při tvorbě databází vzdálených zdrojů, na které odkazují. Lze je tedy použít pro popis elektronických zdrojů off-line i on-line. Nezávislé hlavičky TEI mají stejnou strukturu jako hlavičky TEI. Mají pouze více povinných a doporučených prvků a měly by obsahovat více strukturovaných informací než volného textu.

Hlavičky TEI mohou být uloženy přímo v elektronickém zdroji jako součást kódování zdroje. Vytváří je sám autor dokumentu nebo následně katalogizátor při popisu zdroje. Další možností je uložení záznamu do externí databáze. Hlavičky mohou být různě obsáhlé dle požadavků zpracovatelského subjektu. Autor zdroje může vytvořit zcela jednoduchý popis se základními prvky, kdežto informační instituce může zdroj popsat detailně podle pravidel AACR2 nebo ISBD. Pravidla TEI preferují detailnější popis (Culbertson, 2001; Heery, 1996).

Hlavičky TEI jsou uzavřené mezi párovým tagem <teiHeader>. Skládají se ze čtyř částí, které obsahují informace o souboru, způsobech kódování, věcném profilu textu a historii změn zdroje. Všechny čtyři části jsou dále hierarchicky členěny.

Část *popis souboru* je uzavřena v tagu <fileDesc>. Nese popisné údaje o popisované entitě, ze kterých může být sestavena citace, nebo které může použít knihovník pro vytvoření katalogizačního záznamu. Popis obsahuje také informace o zdrojích, ze kterých je popisovaný zdroj odvozen. Tato část je povinná. Následující tři části jsou volitelné.

V druhé části *popis kódování*, ohraničené tagy <encodingDesc>, jsou informace o tom, zda a jak byl zdroj modifikován vzhledem k původnímu zdroji. Zda došlo ke změně kódování, k transkripci nebo jiným podobným změnám textu.

Profil textu v tagu <profileDesc> obsahuje doplňkové nebibliografické popisné a kontextové informace o textu, např. informace o jazyku zdroje, situaci vzniku. Jsou zde také uvedeny věcné výrazy z různých věcných klasifikací a tezurů.

Poslední část uvozená <revisionDesc> je věnovaná evidenci změn v průběhu vzniku dokumentu nebo jeho modifikací. Tato část je zejména důležitá v případě, kdy je nutné odlišit a identifikovat různé verze jednoho zdroje.

Vzhledem k tomu, že každá z výše uvedených čtyř částí má své dceřinné prvky, může být hlavička TEI ve své maximální podobě velmi obsáhlá. Jak již však bylo zmíněno, povinná je jen její velmi malá část.

Text v prvcích hlavičky TEI můžeme rozdělit do čtyř částí. Základní jsou běžné prvky, které v sobě nesou údaje ve formě volného textu nebo fráze. Prvky, které jsou ukončeny sufixem Stmt, se označují jako shlukující prvky. Obsahují text strukturovaný v dalších dceřinných prvcích. Deklarace se odlišují sufixem Decl a obsahují kódované údaje. Popisné prvky se sufixem Desc obsahují volný text, který může ale nemusí být rozčleněný do dalších dceřinných prvků (Text Encoding Initiative Consortium, 2007a).

Nyní již ale k podrobnějšímu členění čtyř částí hlavičky TEI. Alespoň stručně naznačíme jejich strukturu.

Prvním z nich je popis souboru <fileDesc> s bibliografickými informacemi o zdroji. Obsahuje dceřinné prvky <titleStmt>, <editionStmt>, <extent>, <publicationStmt>, <seriesStmt>, <notesStmt> a <sourceDesc>. Kromě prvku <extent> obsahují všechny další dceřinné prvky.

Prvek <titleStmt> je povinný. Jsou v něm uvedeny informace o názvu díla a intelektuální odpovědnosti. Obsahuje několik volitelných údajů. Kromě názvu se zde vyskytují informace o autorovi či jiných osobách odpovědných za vznik díla:

- název <title> - název zdroje včetně doplňujících názvů a podnázvů, které se uvádějí v opakování prvku
- autor <author> - primární odpovědnost za vznik díla
- ručitel, garant <sponsor> - organizace garantující dílo
- sponzor <funder> - osoba či organizace, která finančně podpořila vznik díla (v rámci projektu)
- ředitel <principal> - vedoucí pracovník, který po odborné stránce ručí za vznik a obsah zdroje
- vedlejší odpovědnost <respStmt> - jiné odpovědnosti za vznik díla
- druh odpovědnosti <resp> - určuje povahu odpovědnosti za dílo
- jméno <name> - vlastní jméno nebo jmenná fráze

Prvek <editionStmt> obsahuje informace spojené s danou verzí textu. Ty jsou uvedené formou volného textu nebo pomocí některého z prvků

- vydání, verze <edition>
- odpovědnost <respStmt>
- jméno <name>
- druh odpovědnosti <resp>

Prvek <extent> udává přibližnou velikost textu pro uložení na přenosné médium. Ta je vyjádřena v některých běžných jednotkách (např. byty, logické jednotky typu věta či odstavec, fyzické jednotky typu disk či CD-ROM)

Další prvek <publicationStmt> je povinný a obsahuje informace týkající se distribuce či vydání textu. Může obsahovat volný text nebo skupinu údajů:

- nakladatel <publisher>
- distributor <distributor>
- majitel autorských práv <authority> - osoba či organizace odpovědná za výrobu kopií zdroje
- místo vydání <pubPlace>
- adresa <address>
- identifikační číslo <idno> - standardní i nestandardní identifikační čísla s určením, o jaký typ čísla se jedná (např. ISBN)
- dostupnost <availability> - status zdroje zahrnuje typy: volný, neznámý, omezený
- datum <date>

Prvek <seriesStmt> zahrnuje informace o příslušnosti k nějaké edici či řadě, pokud existuje. Obsahuje dceřinné prvky:

- název <title> - název edice
- identifikační číslo <idno>
- odpovědnost <respStmt> - vyjádření odpovědnosti za intelektuální obsah edice či řady
- odpovědnost <resp> - druh odpovědnosti
- jméno <name>

Prvek <notesStmt> shrnuje poznámky ke zdroji. Jde např. o původ díla, poznámkové bibliografické údaje, poznámky o souboru a jeho znaky apod. Prvek poznámky <note> sdružuje více poznámek <note>.

Dalším povinným prvkem je <sourceDesc> pro bibliografický popis jednotky, ze které je zdroj odvozen. Může sice obsahovat i volný text, ale doporučuje se použít strukturované údaje:

- citace <bibl> - volně strukturovaná bibliografická citace
- <biblItem> - bibliografická citace obsahující pouze bibliografické dílčí prvky
- <biblFull> - plně strukturovaná bibliografická citace se všemi komponentami popisu
- strukturovaná citace <biblStruct> - přesně strukturovaná citace s údaji ve stanoveném pořadí
- úplná citace <biblFull> - plně strukturovaná bibliografická citace, ve které jsou obsaženy všechny prvky souboru TEI
- seznam bibliografických citací <listBibl>
- pro citaci mluvených projevů lze užít ještě prvky <scriptStmt> a <recordingStmt>

Výše uvedené prvky musí v rámci této sekce následovat v uvedeném pořadí (Text Encoding Initiative Consortium, 2007a).

Druhá část je tvořena prvkem popis kódování <encodingDesc>. Ten opět může obsahovat několik dceřinných prvků. Možné jsou prvky <projectDesc>, <samplingDecl>, <editorialDecl>, <tagsDecl>, <refsDecl>, <classDecl>, <metDecl>, <variantEncoding>. Jak je vidět, jde většinou o prvky s nějakou deklarací.

Prvek <projectDesc> nese informace o důvodu a způsobu vzniku zdroje, kódování apod. Informace jsou vyjádřeny volným textem.

Prvek <samplingDecl> uvádí metody vzorkování textu, metody volby vzorků a velikost zvolených vzorků pro zkoumání, předmět zkoumání apod.

V prvku <editorialDecl> jsou vypsány způsoby a metody, kterými byl text kódován. Obsahuje několik strukturovaných údajů:

- <correction> - udává, za jakých okolností a proč došlo v textu k opravám. Obsahuje další strukturované podúdaje:
 - <status> - údaj o rozsahu opravy textu s možnostmi určení rozsahu oprav a kontrol v textu - velké (high), střední (medium), malé (low), neznámé (unknown)

- <method> - jakým způsobem byly naznačeny opravy ve zdroji. Mezi možnosti patří nenaznačené opravy (silent) a opravy označené dodatky (tags)
- <normalization> - naznačuje míru normalizace použitou při konverzi původního zdroje, zahrnuje zdroj a použité metody
- <quotation> - tento údaj vyjadřuje, jakým způsobem bylo nakládáno při konverzi zdroje s uvozovkami a citacemi, specifikuje, zda byly přeneseny všechny citace, některé nebo všechny citace a jakým způsobem jsou v novém textu citace uvedeny
- <hyphenation> - způsob nakládání a vyjádření spojovníků, čárek a podobné interpunkce
- <segmentation> - způsob dělení textu na věty, odstavce a jiné jednotky
- <stdVals> - specifikace formátů pro zápis data a jiných číselných údajů
- <interpretation> - analytické či vysvětlující informace přidané zdroji v rámci transkripce

Prvek <tagsDecl> obsahuje informace, jak často se daný tag objevuje v textu. Prvky, které je možné v jeho rámci použít jsou:

- <rendition> - informace o zamýšlených vlastnostech pro zobrazení spojených s daným prvkem
- <namespace> - formální název jmenného prostoru, ke kterému prvek patří
- <tagUsage> - užití prvku v daném textu – počet výskytů, identifikační číslo, způsob zobrazení

Prvek <refsDecl> určuje způsob vytváření odkazů pro zdroj pomocí podřízených prvků:

- <refsDecl> - způsob tvorby odkazů
- <cRefPattern> - definuje, jak vytvářet URI
- <state> specifikace součásti odkazu definovaného pomocí milníkové metody. Obsahuje zřetěžené hodnoty, které ohraničují určitý text.

Schémata a třídící kódy, pokud byly použity, jsou definovány v prvku <classDecl>. Určují typologii textu (např. typ zdroje, periodicitu, tematickou příslušnost textu), nebo citují použitá třídící schémata pomocí prvků:

- <classDecl> - specifikuje použité klasifikace v textu
- <taxonomy> - definuje typologii pro určení typu textu

- <category> - obsahuje jednotlivé popisné kategorie textu
- <catDesc> - popisuje kategorii v rámci určité taxonomie nebo textové typologie

Poslední tři prvky části <encodingDes> jsou <fsdDecl> pro určení systému znaků v textu, <metDecl> pro metrické notace a <variantEncoding> určující metody pro jiná kódování. Již neobsahují žádné dceřinné prvky.

Třetí hlavní část hlavičky TEI tvoří prvek věcný profil textu <profileDesc>. Obsahuje tři volitelné komponenty <creation>, <langUsage> a <textClass>. Mohou se v něm objevit v některých případech ale také prvky <textDesc> (údaje pro popis situace vzniku díla), <particDesc> (identifikace zúčastněných rozhovoru) a <settingDesc>.

Prvek <creation> určuje původ textu, datum ap.

Informace o jiném jazyce, nářečí či dialektu, které jsou použité v textu, jsou uloženy v prvku <langUsage>. Prvek obsahuje opakovatelný dceřinný prvek <language> pro popis jednoho nářečí či jazyku s trojznakovým identifikátorem, určením systému písma pro jazyk a přibližným určením míry užití daného jazyka v textu.

Třetí prvek <textClass> třídí zdroj pomocí notací mezinárodních třídění, klíčových slov, vybraného tezauru či jiného třídícího schématu. Používá k tomu prvky <keywords>, <classCode> a pro jiná třídění prvek <catRef>.

Poslední částí hlavičky je prvek historie změn zdroje <revisionDesc> pro zachycení historie změn a editace textu a celé historie souboru. Zajišťuje informace důležité pro odlišení verzí zdroje. Pro vyjádření změn používá prvek <change>, jehož opakování obsahuje jednotlivé změny souboru (Text Encoding Initiative Consortium, 2007a).

Pro popis zdrojů, které vycházejí z mluvených projevů, jsou nutné speciální prvky:

- <scritpStmt> - uvádí citaci scénáře či rukopisu, který byl použit pro řečnický projev
- <recordingStmt> - záznam užitý pro přepis mluvené řeči
- <recording> - zaznamenává detaily týkající se okolností projevu – audio či video záznam, rozhlasové vysílání, živá řeč. Používá údaje jako druh záznamu (<audio>, <video>) a doba trvání záznamu
- <equipment> - technické detaily média či vybavení pro audio či video záznam

- <broadcast> - záznam rozhlasového vysílání

Záznam může obsahovat také volný text. Může nést informaci o tom, kdo záznam mluvené řeči zaznamenal, o odpovědnosti, jménech, datech, místech a dalších údajích, které se vztahují ke vzniku záznamu (Humanities Text Initiative, 1993).

V příloze 6 je uvedena ukázka hlavičky TEI. V ní lze vidět hierarchické vnoření jednotlivých prvků v rámci čtyř hlavních částí hlavičky TEI. Počáteční a ukončovací tagy každé části jsou zvýrazněny tučně stejně jako základní prvek <teiHeader>, který ohraničuje celou hlavičku TEI. V jednotlivých částech je naznačena struktura vnořených prvků.

Hlavičky TEI mají sloužit jako zdroj informací čitelný strojem. Vycházejí z klasického popisu. Vzhledem k nutnosti konverze dat se u některých prvků doporučuje při tvorbě hlavičky TEI dodržovat normy ISBD, AACR2 a americkou národní normu ANSI 39.29 (Humanities Text Initiative, 1993).

5.1.4 Shrnutí informací k formátu TEI

Formát TEI je používán v řadě rozličných projektů a systému. Je možné se o tom přesvědčit na stránce, kde konsorcium TEI udržuje seznam projektů, které formát TEI používají. Kromě projektů, které zaštiťují univerzity a obdobné vzdělávací instituce, je mezi nimi možné najít i projekty komerčních institucí. Jde například o plnotextovou databázi anglické poezie od doby Anglo-Saské do 19. století, kterou s pomocí formátu TEI vytváří a zpřístupňuje Chadwyck-Healey. Databáze obsahuje zhruba 4450 děl (Text Encoding Initiative Consortium, 2001).

Co se týče předmětové orientace souvisejících projektů, převládají projekty zaměřené na literární a lingvistické texty, historické materiály, korespondenci a také na jazykové korpusy. Objevují se ale také projekty orientované na náboženské texty, hudbu nebo filosofii.

5.2 Oxfordský textový archiv



Oxfordský textový archiv (The Oxford Text Archive, OTA, <http://ota.ahds.ac.uk/>) založil v roce 1976 Burnard Lou v rámci programu Oxfordská digitální knihovna (Oxford Digital Library, ODL). OTA je britský digitální archiv elektronických textů a lingvistických korpusů, který slouží akademickým pracovníkům pro výuku a vědecké účely. OTA neprodukuje žádné texty, pouze přebírá a ukládá existující texty (Oxford Text Archive, 2006).

Činnost OTA je sponzorována z několika zdrojů. Hlavní partnerskou organizací je od doby jeho vzniku v roce 1996 Služba pro data v oblasti umění a humanitních vědách (Arts and Humanities Data Service, AHDS), která je sponzorovaná Spojenou komisí pro informační systémy (Joint Information Systems Committee). Dalšími sponzory jsou Středisko výpočetních služeb Oxfordské univerzity (Oxford University Computing Services) a Výzkumná komise pro umění a humanitní vědy (Arts and Humanities Research Board).

OTA poskytuje uživatelům různé služby. Především zajišťuje přístup ke sbírkám elektronických textů, referenčních materiálů a jazykových korpusů. Velká část těchto materiálů je přístupná zdarma. Tyto texty a další ukládané materiály jsou katalogizovány pracovníky OTA. Podrobný popis podle pravidel ISBD(ER) je v metadatovém formátu TEI. Formát TEI slouží zároveň i pro strukturování textu. OTA nabízí také rady a pomoc při vytváření elektronických textů a konzultace při volbě softwaru a standardů pro tvorbu a uchovávání elektronických textů (Oxford Text Archive, 2006).

V současnosti obsahuje fond OTA zhruba 2 500 elektronických textů, slovníků, referenčních materiálů a jazykových korpusů v 26 jazycích. Jde především o literární texty různých žánrů a z různých časových období a materiály k oblasti lingvistiky. Texty jsou vybírány podle profilu fondu. Hodnotí se obsah textu, předpokládané využití textu a také stupeň a kvalita dokumentace¹⁷, která je k textu připojená. Záleží také na formátu, ve kterém je text uložen. OTA se snaží zajistit

¹⁷ Jde o dokumentaci k obsahu textu (hlavička TEI) a informace o formátu a dalších technických parametrech textu.

trvalou archivaci uložených textů a jejich čitelnost. Bere ohled i na možné změny uložení a migraci dat při přechodu na novější technologie. Při výběru textů jsou preferovány texty uložené ve formátu TEI nebo XML, příp. SGML, k archivaci se přijímají také texty ve formátu LaTeX, RTF, nebo prostý text. Formáty typu Word, databáze, tabulky apod. je možné upravovat.. Jeden text může být v archivu uložen ve více formátech.

Autoři, jejichž díla jsou v archivu OTA ukládána, dávají OTA nevýhradní licenci na distribuci textu. OTA má díky této licenci právo kopírovat a zpřístupňovat uložený text. Pokud si to autor přeje, je možné přístup k textu omezit určitými přístupovými právy (Oxford Text Archive, 1998).

5.2.1 Popisné jednotky

Popisnými jednotkami v archivu OTA jsou elektronické texty, jedno i více jazyčné slovníky, databáze, aplikační software, referenční materiály a jazykové korpusy. Jde tedy popis na úrovni jednotky, nejsou popisovány vyšší hierarchie typu sbírka apod. V případě lingvistických korpusů je za jednotku považován celý korpus (Oxford Text Archive, 2006). Pro bibliografické údaje o textu jsou používány hlavičky TEI, které budou tématem následujícího textu.

Zaměření je spíše humanitního směru (lingvistika, historie, právo, ...). Systém neomezuje chronologicky výběr ukládaných zdrojů. Přestože základem je popis elektronických zdrojů, nechybí ani odkazy na tištěné materiály, které se elektronickým zdrojem nějak souvisí (Oxford Text Archive, 1998).

5.2.2 Metadatové záznamy

Preferovaný formát pro popis je TEI lite. Pro popis záznamu o zdroji se používá několik typů metadat. Jako sdílený metadatový formát pro vyhledávání se používá metadatový formát Dublin Core. Jako interní se používá formát TEI Lite, protože popis pomocí tohoto formátu je bohatší a zahrnuje popis zdroje i jednotlivých významových prvků samotného zdroje. Administrativní metadata jsou obvykle obsaženy v hlavičce TEI. Výjimkou jsou informace o uživateli zdroje, které se kvůli důvěrnosti těchto informací uchovávají v samostatné databázi (Oxford Text Archive, 1998).

Pro bibliografický popis textů jsou používány hlavičky TEI. Hlavička TEI může obsahovat čtyři prvky. Jsou to prvky <fileDesc>, <encodingDesc>, <profileDesc> a <revisionDesc>. Kromě prvku <fileDesc> (popis souboru) jsou všechny prvky volitelné a mohou obsahovat další vnořené prvky. Hlavička TEI může být podle potřeby velmi stručná nebo naopak velmi rozsáhlá.

Prvky v hlavičce se dají rozdělit do tří skupin. Prvky s příponou Desc (Descriptions, Popis) nesou popis nějakých vlastností textu a obsahují volný text, nebo další vnořené prvky. Prvky s příponou Stmt (Statements, Přehled) sdružují prvky, které nesou další strukturované informace. Prvky s příponou Decl (Declarations, Prohlášení) obsahují informace o použitém kódování.

Prvek <fileDesc> obsahuje bibliografický popis elektronického textu. V prvku <encodingDesc> jsou údaje o vztahu mezi popisovaným textem a zdrojem, ze kterého byl popisovaný text odvozen. Prvek <profileDesc> nese informace o jazyku textu, důvodu, proč byl text vytvořen apod. Prvek <revisionDesc> obsahuje informace o změnách textu (Morrison, 1998).

Hlavička TEI je uložena v souboru před samotným textem. Využívá se při vyhledávání. Další možné využití metadat uložených v hlavičce TEI je při tvorbě rejstříků, dynamicky generovaných katalogů a vytváření katalogizačních záznamů (např. po převodu do formátu MARC) (Morrison, 1998).

Ukázku hlavičky TEI, která byla vyhledána v Oxfordském textovém archivu, obsahuje příloha 7. V záznamu jsou použity všechny čtyři základní prvky hlavičky TEI. Záznam je v souladu s ISBD. Vzhledem k tomu, že jméno autorky je uvedené včetně data narození a úmrtí, lze soudit, že byl použit autoritní soubor. Z části <encodingDesc> je vidět, že bylo použito třídění OTASH (Oxford Text Archive Subject Headings) a LCSH (Library of Congress Subject Headings). Jasně vidět jsou i zásahy do záznamu, kdo, kdy a co upravoval.

Uvedený záznam se zobrazuje s prvky psanými velkými písmeny. Není to ovšem obecný rys záznamů uložených v OTA. Tento prvek není zřejmě sjednocen, protože u jiných záznamů se zobrazují prvky malými písmeny. Záleží na editoru, ve kterém je záznam vytvářen a na jednotlivých zpracovatelích.

5.2.3 Vyhledávání

Pro vyhledávání v archivu OTA je možné použít jak listování v rejstřících tak vyhledávací formuláře. Pro listování jsou dostupné rejstříky autorů, jazyků a názvů. OTA nabízí vyhledávací formuláře pro jednoduché, pokročilé i fulltextové vyhledávání. Dokonce je možné listovat v tištěném katalogu, který je dostupný ve formátu PDF.

Jednoduché hledání nabízí pouze zadání autora nebo slova či fráze z názvu. Pokročilé vyhledávání dovoluje prohledávat autory, názvy, předmětová hesla a jazyk, spojovat je pomocí boolovských operátorů a používat zkracování výrazů zástupnými znaky. Fulltextové vyhledávání dovoluje vybrat prvek formátu TEI Lite a v něm vyhledat slovo nebo slovní spojení.

5.2.4 Shrnutí informací k archivu OTA

Oxfordský archiv textů je ukázkou fungujícího archivu textů popsaných pomocí hlaviček TEI. Funguje od roku 1976, je tedy možné říct, že má dlouholetou tradici a praxi v popisu elektronických textů. Dne 30. září 2006 oslavil výroční konferencí 30. výročí svého založení. Svůj provoz a podrobnou katalogizaci textů si může dovolit díky tomu, že slouží akademické sféře, která ho také sponzoruje.

5.3 Standard pro kódování korpusů



Standard pro kódování korpusů (Corpus Encoding Standard, CES, <http://www.cs.vassar.edu/CES/>) je výsledkem spolupráce projektu EAGLES¹⁸, projektu Multext¹⁹ a CRNS²⁰. Tyto projekty, zaměřené na kódování jazyka a jazykové aplikace, vznikly v důsledku nekompatibility mnoha stávajících jedno a více jazyčných zdrojů pro sjednocení jejich popisu a vytvoření určitého standardu. Na řešení tohoto problému začala v roce 1994 spolupracovat komunita jazykových organizací - Americké konsorcium pro lingvistická data (U.S. Linguistic Data Consortium), Iniciativa evropského korpusu (European Corpus Initiative, ECI), Britský národní korpus (British National Corpus, BNC), Evropská asociace jazykových zdrojů (European Language Resources Association, ELRA) (Ide, 1996).

Pojem korpus může mít několik významů. V současnosti je pro strojové zpracování korpus vnímán jako „rozsáhlý vnitřně strukturovaný a ucelený soubor textů daného jazyka elektronicky uložený a zpracovávaný“ (Pala, 1996), který slouží k dalšímu jazykovému výzkumu. Lze ho také chápat jako velmi rozsáhlý a složitý soubor textů přirozeného jazyka, jehož vytvoření a další využívání umožňují počítače.

Korpusy jsou vytvářeny tak, že jazyková data jsou ve svém přirozeném kontextu, korpus je natolik obsáhlý, že převažují typické jazykové jevy a data jsou dostatečně reprezentativní, aby na nich bylo možné zakládat studie a analýzy.

¹⁸ Iniciativa Odborná poradní skupina pro standardy lingvistického inženýrství EAGLES (Expert Advisory Group on Language Engineering Standards) spadá pod Evropskou komisi do programu Lingvistický výzkum a inženýrství. Zabývá se výzkumem v oblasti jazykových korpusů, počítačovou lingvistikou, vytvářením lingvistického softwaru apod. Iniciativa je řízena z Itálie z města Pisa (Expert Advisory Group on Language Engineering Standards, 2000).

¹⁹ Projekt Nástroje a korpusy pro vícejazyčný text Multext (Multilingual Text Tools and Corpora) sdružuje skupinu projektů, které vyvíjejí standardy pro kódování jazykových korpusů a nástroje pro jejich zpracování. Multext vyvíjí nástroje pro různé jazyky (bulharština, čeština, angličtina, francouzština, slovenština, svahilština,...). Multext spolupracuje i s dalšími organizacemi a projekty, jejichž výzkum leží v oblasti lingvistiky (Centre National de la Recherche Scientifique, 1996).

²⁰ Francouzské Národní centrum pro vědu a výzkum (Centre National de la Recherche Scientifique, CNRS) je základní veřejnou vědeckou organizací ve Francii. Zabývá se výzkumem v mnoha oborech (nukleární fyzika, fyzika a matematika, komunikace a informační věda, informační technologie, inženýrské obory, vědy o životě, humanitní a sociální vědy). CNRS se snaží podporovat vědu a výzkum a napomáhat vzájemné komunikaci mezi odborníky (Centre National de la Recherche Scientifique, 2006).

Korpusy lze dělit na obecné a speciální. Obecné zachycují jazyk v co nejširším rozsahu a úplnosti. Speciální korpusy se zaměřují na určitou oblast, např. korpus děl určitého autora nebo pouze jednoho díla. Další dělení korpusů je možné z hlediska stáří jazyka. Korpusy synchronní zachycují současný jazyk, korpusy diachronní zachycují jazyk určitého historického období, jsou založeny na vzorcích textů a bývají méně obsáhlé. Podle způsobu komunikace se dělí korpusy na korpusy mluveného jazyka a korpusy psaného jazyka. Tvorba korpusů mluveného jazyka je velmi nákladná, protože je nutné mluvený text přepsat do strojem čitelné podoby. Korpusy psaného jazyka používají existující elektronické texty knih, novin, časopisů apod. Obvykle se takové texty musí upravovat (zbavovat obrázků, konvertovat do jednotného formátu, značkovat a anotovat apod.) (Vassar College, 2002).

Základními nástroji, které pracují s korpusy, jsou obvykle konkordanční programy. Ty třídí a počítají frekvenci slov a slovních spojení v korpusu. Konkordanční programy nerozlišují hononyma (víceznačná slova). Proto se ke korpusu váží také gramatické analyzátoři, které se zaměřují na morfologické, syntaktické a sémantické znaky textů. Přiřazují každému slovu gramatickou značku (slovní druh, gramatická kategorie apod.). Takto zpracovaný korpus se označuje jako anotovaný (Pala, 1996).

CES slouží k popisu všech typů textů. Texty mohou být jednojazyčné i vícejazyčné (Ide, 1996). Texty zpracované ve standardu CES mohou sloužit i pro kvantitativní analýzy dat (frekvence slov, znaků) i pro kvalitativní analýzy dat (gramatika, zjišťování mutací slov, generování konkordancí). Pro analýzy jsou důležité lingvistické entity - části textu dělené na větší celky (odstavce, kapitoly) a menší části (věty, jména, zkratky, jednotlivá slova).

CES byl navržen pro optimální využití v oblasti jazykového inženýrství a výzkumu jako mezinárodně uznávaný kódovací standard pro kódování aplikací jazykových korpusů v přirozeném jazyce. CES je založen na formátu TEI a vychází i ze SGML a XML. CES nepoužívá celý formát TEI. Používá vybrané části, které jsou vhodné pro kódování jazykových korpusů. Naopak pro jazykové oblasti, které formát TEI neřeší, CES formát TEI rozšířil. Všechny změny, které CES ve formátu TEI udělal, byly předány přímo iniciativě TEI jako podněty pro další rozvoj formátu TEI (Ide, 1996).

CES stanoví minimální kódování pro popis (strukturní a typografické údaje) a obecnou architekturu kódování tak, aby vyhovovala struktuře textových databází. Určuje také rozšířené kódování včetně lingvistických poznámek (Ide, 1996).

Pro popis korpusu pomocí CES v jazyce XML vytváří CES formát XCES (Corpus Encoding Standard for XML). XCES využívá také možností XML v oblasti odkazování pomocí technologií XLink a XPointer (Vassar College, 2002).

5.3.1 Popisné jednotky

CES popisuje lingvistický korpus jednojazyčný nebo vícejazyčný. Korpusem může být určitý text (poezie, drama, mluvený projev, ...). Obecně lze za korpus považovat jakýkoliv soubor lingvistických dat. CES se zaměřuje na lingvistické objekty, za které považuje nejen celé texty, ale i jejich části. Jde např. o kapitoly či odstavce textu, i o menší části (např. věty nebo jednotlivá slova). CES zachycuje počty jejich výskytů, vztahy lingvistických jednotek, jejich pořadí a vzdálenost apod. a dovoluje nad těmito údaji provádět různé operace, tvořit statistiky a analýzy (Ide, 1996).

Podobně jako TEI, slouží i CES k popisu celého textu. Kóduje pomocí značek jednotlivé části textu, odstavce, slovní spojení, slova. Pro potřeby popisu korpusů byl formát TEI upraven. Některé prvky mají ve formátu CES před jménem prefix „ces“ místo „tei“. Následující text bude věnován hlavičkám CES <cesHeader> (původně <teiHeader>) (Ide, 1996).

Hlavičky CES popisují celý korpus i jednotlivé texty, ze kterých je složen. Celý korpus má hlavičku CES označovanou jako „hlavička korpusu“, jednotlivé části (texty) mají hlavičku CES označovanou jako „hlavička textu“. Oba typy hlaviček jsou sdruženy v prvku <cesHeader> a rozlišeny v rámci tohoto prvku pomocí kvalifikátoru typu (hlavička korpusu nebo hlavička textu). Je možné popisovat více textů dohromady (např. antologie). Popis může být uplatněn na několika úrovních (Vassar College, 2002).

5.3.2 Metadatové záznamy

Součástí popisu CES je hlavička TEI (resp. CES). I v CES nese hlavička nejen bibliografické údaje (název, autor,...) ale i údaje o kódování. Hlavička CES obsahuje prvky vnořené na několika úrovních. Základní prvek <cesHeader> obsahuje

údaje týkající se typu zdroje, který hlavička popisuje (korpus nebo text), jméno autora hlavičky CES, verze hlavičky, její stav (nová, aktualizovaná), datum vytvoření a aktualizace. Do hlavičky CES mohou být vnořeny čtyři prvky: <fileDesc>, <encodingDesc>, <profileDesc> a <revisionDesc>. Vnořené prvky hlavičky CES obsahují další prvky na nižší úrovni. Úroveň vnoření může být více.

Prvek <fileDesc> obsahuje bibliografický popis korpusu nebo jeho části. V tomto prvku mohou být vnořeny prvky <titleStmt>, <editionStmt>, <extent>, <publicationStmt> a <sourceDesc>. V prvku <titleStmt> je název korpusu nebo názvy textů, ze kterých je korpus složen. Prvek <editionStmt> obsahuje informace o dané verzi korpusu či textu, který je v korpusu obsažen. Prvek <extent> zajišťuje informace o velikosti a rozsahu textu. Prvek <publicationStmt> zahrnuje informace o vydání a distribuci korpusu. V prvku <sourceDesc> je vnořen základní bibliografický popis textu či textů, ze kterých byl korpus vytvořen.

Druhou částí hlavičky CES je prvek <encodingDesc>, který udává vztahy mezi korpusem a textem nebo texty, ze kterých byl korpus odvozen. Tento prvek může obsahovat šest prvků na nižší úrovni. Jsou to prvky <projectDesc>, <samplingDecl>, <editorialDecl>, <tagsDecl>, <refsDecl> a <classDecl>. Údaje o projektu, v jehož rámci korpus vznikl, jsou uvedeny v prvku <projectDesc>. Poznámka ke způsobu vytvoření korpusu a metodě výběru textů pro korpus je uvedena v prvku <samplingDecl>. V prvku <editorialDecl> jsou další prvky, které specifikují způsoby zpracování jednotlivých částí korpusu. Prvek <tagsDecl> se používá jiným způsobem v hlavičce korpusu a jinak v hlavičce textu. V hlavičce korpusu se používá k vyjmenování všech prvků, které jsou v korpusu použity a jejich krátkému popisu. V hlavičce textu se používá k vyjmenování použitých prvků a počtu jejich použití v textu. Každý prvek je vložen ve značce <tagUsage>, pojmenován a určen počet výskytů. Prvek <refsDecl> obsahuje informace o způsobech odkazování v původních textech, které byly použity při tvorbě korpusu. V prvku <classDecl> jsou údaje o kategoriích textů v korpusu.

Třetí prvek hlavičky CES <profileDesc> nese další informace o textu a jeho vzniku (jazyk, kontext vzniku textu, třídění,...). Obsahuje prvky <creation>, <langUsage>, <wsdUsage>, <textClass>, <translations> a <annotations>. Prvek <creation> obsahuje informace o původu textu. Informace o jazycích, dialektech a dalších jazykových skupinách, které jsou obsaženy v korpusu nebo textu, jsou v prvku <langUsage>. Každý jazyk má určený identifikátor a znakovou sadu. Prvek

<wsdUsage> obsahuje informace o všech znakových sadách, které jsou v korpusu použity. Prvek <textClass> používá k vyjádření předmětu a témat v textech tezaury a třídění. V prvku <translations> jsou informace o existujících překladech textů. Prvek <annotations> sdružuje informace o existujících souborech anotací, které se váží k textům.

Prvek <revisionDesc> je čtvrtým prvkem hlavičky CES. Shrnuje historii editací a úprav textu. Jednotlivé změny jsou uloženy v prvku <change>, který je prvku <revisionDesc> podřízen (Ide, 1996).

Hlavička XCES v XML <xcesHeader> může být vnořena ve zdroji nebo může být uložena v externím souboru a provázána se zdrojem pomocí odkazu. Hlavička XCES vychází z hlavičky CES. Obsahuje několik dalších prvků.

V části <profileDesc> přibyl prvek <particDesc> pro popis účastníků lingvistického projevu. Prvek <person> určuje jednotlivce, kteří se účastní lingvistického projevu. Obsahuje identifikátor osoby, údaje o pohlaví a věku osoby. Je-li osob více, každá má vlastní prvek <person>. Více prvků <person> je sdruženo pomocí prvku <personGrp>. Aby bylo možné mezi jednotlivými osobami definovat vztahy, byl definován prvek <particLinks> a jeho vnořený prvek <relation>. Částečnou strukturu prvku <profileDesc> ukazuje příklad 11. Popisuje ženu ve věku 42 let, jejíž identifikátor je „p1“. Krátký popis uvádí její vzdělání, místo a datum narození atd. Druhou osobou je muž ve věku 43 let. Jeho identifikátor je „p2“. V prvku <particLinks> je uvedeno, že tyto osoby jsou manželé (Vassar College, 2002).

Příklad 11: Ukázka použití prvku <particDesc>

```
<particDesc>
<person id="p1" sex="f" age="42">Female informant, well educated, born in Boston US, 12 Jan 1950, of unknown occupation.
Speaks English fluently.</person>
<person id="p2" sex="m" age="43"/>
<particLinks>
<relation active="p1 p2" desc="spouse"/>
<particLinks>
</particDesc>
```


Dalším prvkem, který byl přidán do prvku <profileDesc> v hlavičce XCES je prvek <settingDesc>. Tento prvek nese údaje o okolnostech, které provázely lingvistický projev. Pro jednotlivé okolnosti slouží prvek <setting>, který je dále rozvíjen pomocí prvků <name> (jméno události, místa či jiné okolnosti), <time> (pro určení časového období projevu) a <locate> (popis místa) (Vassar College, 2002).

Hlavička CES je uvedena v příloze 8 (Ide, 1996). Obsahuje tři základní prvky s několika vnořenými prvky. Základní prvky obsažené v příkladu jsou <fileDesc>, <encodingdesc> a <profiledesc>. Čtvrtý prvek <revisionDesc> není použit. Zajímavý je prvek <wordcount>, který udává počet slov v kódovaném textu. Prvek <tagsdecl> obsahuje prvky, které jsou v textu použity a počet jejich výskytů. Záznam ukazuje také dceřinné prvky prvku <editorialDecl>, které specifikují způsoby zpracování jednotlivých částí korpusu.. Uveden je prvek <correction>, který určuje rozsah úprav, jež byly na původním textu provedeny. Další prvek <segmentation> zase např. vypovídá o způsobech dělení textu na menší jednotky. V případě ukázky jde o popis jednoho kódovaného textu, nikoliv o popis celého korpusu.

Upravená hlavička CES zajišťuje popis celého korpusu i jeho částí. Hlavička CES je v podstatě podmnožinou hlavičky TEI. Byly provedeny následující změny:

- k existujícím prvkům byly přiřazeny atributy specifikující vlastnosti prvků
- prvky a vlastnosti mají omezené hodnoty pomocí kvalifikátorů
- byl zjednodušen obsahový model prvků, prvky mohou obsahovat vnořené prvky nebo volný text (Ide, 1996)
- byly přidány prvky pro vícejazyčné anotace a bibliografické poznámky (Ide, 2002)

5.3.3 Shrnutí informací k formátu CES

Celou strukturou DTD CES a XCES zdroje nemá význam se v této práci zabývat. V podstatě je stejná jako formát TEI, pouze má jednodušší strukturu a je tvořena jen vybranými částmi formátu TEI, které jsou potřebné pro popis lingvistických objektů.

Standard CES a XCES používají lingvistické projekty a instituce pro sjednocení popisu zdrojů. Jde o celou řadu projektů, z nichž můžeme jmenovat např. projekt EAGLES a Multext, dále PAROLE Corpora²¹, TIPSTER²², MATE²³ a další (Ide, 2002).

²¹ Předběžné přípravy pro organizaci lingvistických zdrojů a jazykové inženýrství (Preparatory Action for Linguistic Resources Organisation for Language Engineering, PAROLE) je projekt sponzorovaný Evropskou unií. Vytváří všeobecný slovník a korpusy v jazycích, kterými se mluví v Evropské unii (University of Sheffield, 2001).

²² TIPSTER je repozitář jazykových korpusů popsaných pomocí standardu CES. TIPSTER popisuje texty pomocí architektury, která dovoluje připojovat k částem textu jazykové prvky (jména a data) a strukturní prvky (záhlaví, odstavce). Pomocí těchto prvků mohou být určité části textu blíže specifikovány a mohou k nim být připojeny poznámky (Ogden, 2004).

²³ Projekt Víceúrovňové anotace, nástroje pro inženýrství (Multilevel Annotation, Tools Engineering, MATE) se zaměřuje na opakované využívání jazykových korpusů, jejich vytváření a udržování. Zaměřuje se především na zachycení mluveného textu (MATE, 1998).

5.4 Popis rukopisů a historických fondů

5.4.1 Projekt MASTER

V letech 1999 až 2001 spolupracovalo několik knihoven na projektu MASTER (Manuscript Access through Standards for Electronic Records, Přístup k rukopisům prostřednictvím standardů pro elektronické záznamy, domovská stránka projektu <http://www.cta.dmu.ac.uk/projects/master/index.html> byla zrušena). V té době docházelo k nárůstu digitalizace rukopisů a starých tisků a ukázala se potřeba sjednotit popis těchto dokumentů a vyřešit jejich zpřístupnění v počítačových systémech a sítích. Cílem projektu bylo vytvořit prototyp sdíleného online katalogu středověkých rukopisů latinského písma, který by se stal standardem pro popis rukopisů a pro výměnu a sdílení jejich záznamů (Uhlíř, 1999). Sponzorován byl Evropskou unií v rámci IV. rámcového programu *Telematika pro knihovny* (Telematics for Libraries). Garantem projektu byla Monfortská univerzita v Leicesteru ve Velké Británii, Oxfordská univerzita a mezi účastníky patřila např. Královská knihovna v Hagu, Oxfordská univerzita a také Národní knihovna České republiky. Projekt byl úzce propojen s dalšími projekty, z nichž jmenujme alespoň severoamerické projekty EAMMS (Electronic Access to Medieval Manuscripts, Elektronický přístup ke středověkým rukopisům) a *Digitální písárská dílna* (Digital Scriptorium).

Jako vhodný formát pro zápis záznamů o rukopisech a starých tiscích se ukázal metadatový formát TEI. Na základě tohoto formátu bylo nadefinováno a v červnu 2001 předloženo DTD MASTER pro rukopisy. Na jeho vytvoření pracovali kromě členů projektu MASTER také zástupci TEI (Burnard, 1999; MASTER, 1999; Cover, 2001). Formát byl upraven tak, aby lépe vyhovoval popisu rukopisů (MASTER, 1999).

Vývoj TEI pro popis rukopisů neustal. I v současnosti působí v rámci konsorcia TEI skupina, která se zabývá aplikací TEI na jejich popis. Popis rukopisů se stal součástí TEI P5, konkrétně se jím zabývá část 9 (Sperberg-McQueen, 2007).

5.4.1.1 Bibliografické jednotky

Standard MASTER a z něho i následně vycházející popis ve formátu TEI, byl a je primárně určen pro popis rukopisů. Je ale natolik obecně definován, že může být použit pro popis jakéhokoliv jiného materiálu, např. starých tisků a archiválií. Sám katalogizátor si volí, na jaké úrovni bude materiál popisovat (Sperberg-McQueen, 2007).

Odlišností v případě popisu rukopisů oproti klasickému knihovnickému popisu je také fakt, že knihovnictví chápe jako popisnou jednotku určité vydání. V případě rukopisů je každý rukopis jedinečný a spolu s dalšími rukopisy, příp. jinými historickými zdroji, tvoří soubor pramenů k určitému tématu. Rukopisům, jejich proměnám, vazbám a vzájemným vztahům se věnuje kodikologie (Uhlíř, 1999).

Standard MASTER i TEI pro rukopisy umožňuje popis strukturovat, takže je v rámci popisu jednoho rukopisu možné popsat různé části popisovaného rukopisu a naznačit mezi nimi hierarchické vztahy (Uhlíř, 2002).

5.4.1.2 Metadatové záznamy

Standard MASTER (a v současnosti formát TEI pro popis rukopisů) je uzpůsoben tak, aby vyhovoval specifickému popisu rukopisných dokumentů. Díky možnosti vnořování a opakování jednotlivých prvků je možné informace strukturovat, v rámci jednoho popisu zaznamenat informace o jednotlivých částech, které jednotku tvoří, hierarchizovat i vztahy mezi jednotlivými částmi apod. (Uhlíř, 2002). Základem bibliografického popisu je opět hlavička.

V původním standardu MASTER byl definován prvek <msDescription> a jeho možné použití kdekoli v textu nebo v hlavičce dokumentu. Prvek mohl obsahovat několik dceřinných prvků, které mohly být dále strukturovány. Šlo o prvky <msIdentifier>, <msHeading>, <msSummary>, <msContents>, <physDesc>, <history> a <adminInfo>. Povinný byl pouze <msIdentifier>.

Jediný povinný prvek <msIdentifier> slouží pro jedinečnou identifikaci popisovaného rukopisu v rámci daného repozitáře. Může obsahovat další dceřinné prvky pro určení země, regionu, instituce a další specifikaci uložení dokumentu, jak je vidět na příkladu 12.

Příklad 12: Prvek <msIdentifier> v hlavičce MASTER

```
<msIdentifier>
  <country>UK</country>
  <region>Oxfordshire</region>
  <settlement>Oxford</settlement>
  <institution>The University of Oxford</institution>
  <repository>Bodleian Library</repository>
  <collection>Rawlinson</collection>
  <idNo type="BOD">Rawl. MS poet. 176</idNo>
  <altName type="nick">The Tungebryht manuscript</altName>
</msIdentifier>
```

Prvek <msHeading> obsahuje stručné záhlaví nebo katalogizátorem dodaný název.

Minimální katalogizační údaje o rukopisu zapsané ve speciálním formátu jsou uloženy v prvku <msSummary>, který opět obsahuje několik samostatných prvků. Obsah těchto prvků by měl být formalizován, neměl by obsahovat volný text, ale měl by se blížit popisu podle katalogizačních pravidel. Příklad 13 ukazuje použití tohoto prvku.

Příklad 13: Prvek <msSummary> v hlavičce MASTER

```
<msSummary>
  <author>Domenico Cavalca</author>
  <title>Vite dei santi padri</title>
  <origPlace>Naples</origPlace>
  <origDate>1474</origDate>
</msSummary>
```

Část <msContents> popisuje z hlediska intelektuálního obsahu jednotlivé části rukopisu nebo starého tisku. Volba, jaká část textu bude považována za samostatnou intelektuální část, je ponechána na katalogizátorovi. Samostatně může být popsáno dílo určitého autora, pokud se dokument skládá z více děl různých autorů, ale samostatně lze popsat také jednotlivou tabulku, kapitolu nebo báseň v díle. Pro samostatně popsané části se používá opakování prvku <msItem> uvnitř prvku <msContents>.

Všechny informace související s fyzickými aspekty rukopisu (formát, materiál, písmo, číslování listů, vazba, ...) jsou uloženy v prvku <physDesc>. Prvek opět může obsahovat několik samostatných dceřinných prvků, např. <binding>, <music>, <collation>.

Prvek <history> může obsahovat informace o původu, provenienci a akvizici rukopisu nebo starého tisku. Pro každou z těchto informací slouží vlastní dceřinný prvek <origin>, <provenance> a <acquisition>.

Prvky <recordHist>, <custHist>, <availability> sdružuje prvek <adminInfo>. Obsahují stručné informace vztahující se ke správě a kuraci rukopisu nebo starého tisku. Prvek <recordHist> obsahuje údaje o změnách samotného metadatového záznamu a může mít dceřinné prvky <source> a <change>.

V popisu je možné zachytit odkazy na dokumenty odvozené z rukopisu nebo starého tisku (kopie, i elektronické), případně na části rukopisu nebo starého tisku, kterou nejsou jeho plnou součástí, ale jsou k němu nějak přidružené. Je k tomu možné použít speciální prvek <surrogates>, nebo, pro části dokumentu přidružené k dokumentu, prvek <accMat> (Burnard, 1999, 2001).

Obr. 8: Formulář programu Xmetal pro katalogizaci záznamů ve formátu MASTER

The screenshot shows the Xmetal program interface with a menu bar (File, Edit, View, Insert, Format, Tools, Table, Window, Master, Help) and a toolbar. The main window displays a form for manuscript description. At the top, there's a list of languages with checkboxes, including Khasa (kha), Khotanese (kho), Konkani (kon), Kurukh (kru), Lahnda (lah), Lushai (lus), Magahi (mag), Maithili (mai), Malayalam (mal), Manipuri (mni), Marathi (mar), Marwari (mwr), Munda languages (mun), Nepali (nep), Newari (new), Oriya (ori), Pahlavi (pal), Pali (pli), Panjabi (pan), Persian (per), Prakrit languages (pra), Pushto (pus), Rajasthani (raj), Sanskrit (san), Santali (sat), Sindhi (snd), Sinhalese (sin), Sino-Tibetan (Other) (sit), Tamil (tam), Telugu (tel), Thai (tha), Tibetan (tib), and Urdu (urd). Below this, there's a 'Manuscript Description' section with fields for 'msIdentifier', 'Country: United Kingdom', 'Region: Oxfordshire', 'Settlement: Oxford', 'Institution: University of Oxford', 'Repository: Bodleian Library', 'Collection: Indian Institute', and 'Identification number: 1234Testing'. The bottom status bar shows 'TEI.2 / text / body / msDescription / msIdentifier'.

Pro katalogizaci lze využít formuláře programu Xmetal, které byly vyvinuté v rámci souvisejícího projektu (Jain Cataloging Project) v knihovně Bodleiana. Obrázek 8 ukazuje příklad jednoho formuláře, ve kterém jsou vidět značky pro jednotlivé popisné prvky (Text Encoding Initiative Consortium, 2003).

Současný formát TEI pro popis rukopisů je modifikovaný. Opět definuje prvek <msDesc> (manuscript description), který se používá pro popis jednoho rukopisu. V něm jsou vnořeny dceřinné prvky <msIdentifier>, <head>, <msContents>, <physDesc>, <history>, <additional> a <msPart>.

Povinný je opět pouze prvek <msIdentifier>, který jednoznačně identifikuje popisovaný rukopis. Prvek <head> obsahuje jakýkoliv typ záhlaví, např. název části. Pro popis intelektuálního obsahu rukopisu nebo jeho části slouží prvek <msContents>. Prvek <physDesc> je určen pro plný fyzický popis rukopisu nebo jeho části. V prvku <history> lze popsat historii rukopisu nebo jeho části. Prvek <additional> sdružuje další informace o rukopisu nebo souvisejících opisech. Může obsahovat informace jak bibliografické tak administrativní nebo správní informace. Prvek <msPart> nese informace jednotlivých částech, které byly původně samostatné a nyní tvoří popisovaný rukopis. Jednotlivé popsané prvky mají opět své dceřinné prvky pro detailnější popis.

Příklad 14 ukazuje prvek <msDesc> v popisu rukopisu ve formátu TEI s použitím několika výše jmenovaných prvků (Sperberg-McQueen, 2007).

Původní format MASTER pro popis rukopisů byl zapracován do metadatového formátu TEI. Konsorcium se bude i nadále věnovat otázkám spojeným s touto specifickou oblastí popisu (Sperberg-McQueen, 2007).

Příklad 14: Popis rukopisy ve formátu TEI

```
<msDesc xml:id="MSRP149">
  <msIdentifier>
    <settlement>Oxford</settlement>
    <repository>Bodleian Library</repository>
    <idno>MS. Rawlinson poet. 149</idno>
  </msIdentifier>
  <msContents>
    <msItem defective="true">
      <author>Geoffrey Chaucer</author>
      <title type="uniform">The Canterbury Tales</title>
      <note>Mutilated at beginning and end; contains A431-I1092 in the Riverside edition numbering.</note>
    </msItem>
  </msContents>
  <physDesc>
    <objectDesc form="codex">
      <supportDesc material="perg">
        <support>
          <p>Parchment.</p>
        </support>
        <extent>136 folios: <measureGrp unit="mm">
          <height extent="280"/>
          <width>190</width>
        </measureGrp>
        <collation>
          <p>Twenty-three quires of eight, of which only nine are not defective, and a final (defective) quire of six.</p>
        </collation>
      </supportDesc>
      <layoutDesc>
        <layout columns="1" ruledLines="38 74">
          <p>Margined and ruled with crayon through fol. 51v, thereafter in drypoint. Single columns of 38-74 lines per page.</p>
        </layout>
      </layoutDesc>
    </objectDesc>
    <handDesc hands="4">
      <p>Four hands, varying between <term type="script">cursive anglicana</term> (the first hand) and <term type="script">mixed secretary</term> (the fourth hand). Hand 4 is responsible for around two-thirds of the manuscript, from fol. 45 to the end; hand 3 writes only three lines and a few words on fol. 38r; the other two hands divide the remainder of the manuscript, up to fol. 45, between them.</p>
    </handDesc>
  </physDesc>
  <history>
    <origin>
      <p>Written in <origDate notBefore="1450" notAfter="1475">the third quarter of the 15th century</origDate> in <origPlace>Pembrokeshire, Wales</origPlace>.</p>
    </origin>
  </history>
</msDesc>
```


5.4.2 Projekt MEMORIA



Do projektu MASTER byli zapojeni také zástupci Národní knihovny České republiky. V rámci této spolupráce se rozběhla digitalizace rukopisů. Následovaly aktivity související s iniciativou UNESCO *Paměť světa* a český program *Memoriae Mundi Series Bohemica* (MMSB), který probíhal v rámci programu VISK 6 - *Národní program digitálního zpřístupnění vzácných dokumentů*. To vše se stalo základem pro český projekt MEMORIA a databázi Manuscriptorium (http://www.manuscriptorium.com/Site/CZE/default_cze.asp). Pro popis digitalizovaných rukopisů byl využit právě formát MASTER.

Projekt MEMORIA probíhal v letech 1992 až 1996 a si kladl za cíl vybudovat virtuální badatelské prostředí pro oblast historických fondů, tedy nikoliv pouze pro rukopisy. V současnosti se označení používá pro iniciativu, která se kolem řešení digitalizace, popisu a zpřístupnění starých fondů sdružila. Další aktivity v rámci MEMORIA národního i mezinárodního charakteru jsou podporovány dotacemi z grantů a také financovány ze sponzorských darů.

Manuscriptorium se skládá z Otevřeného katalogu historických fondů a digitalizovaných kopií a případně i plných textů rukopisů, inkunabulí, starých tisků, map a dalších historických materiálů. Otevřený katalog historických fondů je volně dostupný bez omezení. Obsahuje popisné metadatové záznamy vzniklé v souvislosti s digitalizací v projektu MASTER i v dalších návazných projektech (Národní knihovna, 2005). Záznamy a digitalizované dokumenty jsou propojeny. Nejsou však digitalizovány všechny dokumenty, ke kterým existuje v databázi záznam.

Do spolupráce na této databázi se postupně zapojily nejen české ale i zahraniční instituce, např. z Rakouska, Litvy, Německa, Maďarska (Uhlíř, 2005). Digitalizované dokumenty jsou přístupné na základě licenční politiky Národní knihovny a podle vůle jednotlivých spolupracujících partnerů (Národní knihovna, 2006a). Následující text se bude věnovat metadatovým záznamům v bázi Manuscriptoria, nebude zmiňovat formáty pro uchovávání obrazových informací.

5.4.2.1 Popisné jednotky

Popisnými jednotkami jsou jednotlivé rukopisy, případně staré tisky, mapy a jiné historické dokumenty. Do Manuscriptoria patří dokumenty vydané do r. 1800. Záleží na jednotlivých partnerech, na které dokumenty kladou při zpracování důraz.

Podle doporučení Národní knihovny České republiky má nejdříve proběhnout popis dokumentu a následně jeho digitalizace. Popis musí být natolik podrobný, aby umožnil následné propojení s digitalizovaným dokumentem, případně jeho určitými částmi. Formát MASTER tyto předpoklady splňuje, protože dovoluje vytvořit jak jednoduchý popis tak popis jdoucí do hloubky. Navíc formát MASTER umožňuje integrovat data od různých původců (Uhlíř, 2006).

5.4.2.2 Metadatové záznamy

Pro metadatové záznamy byl původně používán formát DOBM²⁴. Byl ale nahrazen formátem MASTER a dříve vytvořené záznamy byly v roce 2003 konvertovány (Psohlavec, 2004). V bázi jsou záznamy uloženy ve formátu XML ve struktuře odpovídající standardu METS. Popisná metadata jsou tedy vytvořena podle standardu MASTER, resp. TEI pro rukopisy (Národní knihovna, 2006a). Používá se pro popis rukopisů i pro popis starých tisků. Je zajištěna podpora konverze záznamů mezi formátem MASTER a formáty typu MARC. Záznamy je možné sdílet přes protokol Z39.50 i OAI-PMH (Uhlíř, 2005).

Aby bylo možné propojovat záznam s digitalizovaným dokumentem, byl formát MASTER rozšířen a vznikl tzv. MASTER+. Pomocí nástroje M-Edit se vytvoří šablona dokumentu s číslovanými úrovněmi pro jednotlivé stránky digitalizovaného dokumentu a následně se nástrojem M-Tool vytvoří linky od číslovaných úrovní stránek k jejich digitálním obrazům (Uhlíř, 2006).

Protože povinné prvky, tak jak je stanoví formát MASTER, nebyly pro identifikaci a ukládání dokumentů a metadat do databáze plně dostačující, stanovila Národní knihovna určité prvky pro minimální záznam. Pokud informace existují, jsou prvky povinné.

²⁴ Formát DOBM DOBM (Digitised Old Books and Manuscripts) vytvořila Národní knihovna ve spolupráci s firmou AiP Safe pro popis elektronických dokumentů v rámci projektu *Paměť světa* (Mottl, 2001).

V prvku msDescription / msIdentifier jde o dceřinné prvky:

- místo uložení (settlement) – město nebo jiná místní identifikace uložení dokumentu
- knihovna (repository) – instituce, kde je dokument uložen
- signatura (idNo) – signatura nebo jiná identifikace dokumentu

V prvku msDescription / msHeading jsou povinné prvky:

- hlavní název (title) – název dokumentu, případně souborné označení více textů v jednom dokumentu
- autor (author) – autor nebo více autorů daného textu z hlediska intelektuálního původu textu
- rok vydání (origDate) – doba vzniku dokumentu
- jazyk originálu (textLang) – jazyk textu, příp. více jazyků
- poznámka (note) – libovolné poznámkové informace o dokumentu

Prvek msDescription / msContents by měl obsahovat minimálně prvek obsah (overview) s krátkým shrnutím obsahu dokumentu.

Další částí je msDescription / physDesc s doporučenými prvky:

- iluminace (decoration / decoNote) – informace o výzdobě dokumentu
- notace (musicNotation) – informace o notaci, je-li v textu obsažena
- vazba (bindingDesc / binding) – popis vazby dokumentu
- materiál (support)
- rozsah (extent) – počet stran včetně předních a zadních předsádek
- rozměry (extent) – možnost zadat rozměry jednotlivých listů

Záznam by měl obsahovat také část msDescription / additional minimálně s prvkem literatura (listBibl / bibl), ve kterém jsou odkazy na literaturu vztahující se k popisovanému dokumentu. V záznamu je možné použít také část s údaji o historii. Pro popis starých tisků je formát dále rozšířen o prvky pro místo tisku, jméno tiskaře, místo vydání a jméno vydavatele (Kučera, 2006).

5.4.2.3 Vyhledávání

Metadatové záznamy jsou uloženy ve volně dostupném Otevřeném katalogu historických fondů, který tvoří základní prvek Manuscriptoria. Pro vyhledávání je k dispozici rozhraní jednoduchého i pokročilého vyhledávání s rejstříky. Příklad 15 obsahuje jeden z kratších vyhledaných záznamů. Obsahuje základní části popisu uzavřené v prvku <msDescription>.

Příklad 15: Záznam ve formátu MASTER vyhledaný v databázi Manuscriptoria

```
<msDescription lang="CZ">
<msIdentifier>
<settlement>Brno, Smetanova 14</settlement>
<repository>Moravské zemské muzeum, oddělení dějin hudby MZM, Smetanova 14, 602 00 Brno </repository>
<idno>A 8.762</idno>
</msIdentifier>
<msHeading>
<title>Varhanní tabulatura</title>
<author>Anonymi</author>
<origDate>1620-1680</origDate>
<textLang/>
<note>Pochází ze starobrněnského kláštera.</note>
</msHeading>
<msContents>
<overview><p>Skladby pro varhany.</p></overview>
<msItem/>
</msContents>
<physDesc>
<support><p>papír, kůže</p></support>
<extent>86 listů (vpředu i vzadu jedna mladší předsádka) <dimensions>150 x 195 mm</dimensions></extent>
<musicNotation><p>Varhanní tabulatura o dvou osnovách, šestilinkové pro pravou a sedmilinkové pro levou ruku.</p></musicNotation>
<decoration>
<decoNote><p></p></decoNote>
</decoration>
<bindingDesc>
<binding><p>nová kožená vazba</p></binding>
</bindingDesc>
</physDesc>
<additional>
<listBibl>
<bibl>Theodora Straková: Anonymní varhanní tabulatura moravského původu. ČMM 35/1950, s. 439-455.</bibl>
<bibl>Theodora Straková: Starobrněnská varhanní tabulatura. Musikologie 5/1958, s. 8-48.</bibl>
<bibl>Leoš Janáček: Starý rukopis notový z archivu Králové kláštera na St. Brně. Hudební listy IV/1887-8, s. 1 a d.</bibl>
</listBibl>
</additional>
</msDescription>
```

Manuscriptorium se stalo spolu s výstupy programu Kramerius²⁵ a projektu WebArchiv základem Národní digitální knihovny (<http://www.ndk.cz/>), která má zajistit trvalé uchování národního kulturního dědictví. Od 1. září 2006 je na základě dohody s Ministerstvem školství, mládeže a tělovýchovy o hromadné licenci pro studijní účely povolen přístup pro všechny české školy.

²⁵ Program Kramerius se zaměřuje na digitalizaci novodobých dokumentů.

5.4.3 Shrnutí informací k popisu rukopisů

Popis rukopisů a celkově historických fondů je významnou složkou uchovávání a zpřístupňování kulturního bohatství. Popis historických fondů řeší své specifické otázky spojené s určitými odlišnostmi historických fondů od zdrojů běžně popisovaných v knihovnické praxi. Tyto odlišnosti byly zohledněny v projektu MASTER při vytváření DTD pro popis rukopisů. Původní format MASTER se stal součástí metadatového formátu TEI a je předpoklad, že jeho rozvoj bude konsorcium TEI koordinovat i nadále.

Na základě českého projektu MEMORIA vznikla iniciativa MEMORIA, která vybudovala databázi historických fondů Manuscriptorium. Metadatové záznamy jsou dostupné volně a digitalizované dokumenty a jejich plné texty dle licenčních podmínek. Vybrané dokumenty ze sbírky jsou přístupné široké veřejnosti. Databáze se stále rozvíjí a stoupá její využití. Do databáze přispívá řada českých knihoven a dalších kulturních institucí (Moravská zemská knihovna v Brně, Státní technická knihovna Praha, Památník národního písemnictví, Královská kanonie premonstrátů na Strahově, Klášter premonstrátů Teplá, Arcibiskupský zámek a zahrady Kroměříž, Vlastivědné muzeum a galerie v České Lípě,...) a zapojují se také zahraniční instituce (Library of Lithuanian Academy of Sciences, Biblioteka Uniwersytecka we Wrocławiu, University Library Budapest, Univerzitná knižnica Bratislava, Ústredná knižnica Slovenskej akadémie vied) (Národní knihovna, 2006b).

Věřím, že aktivity související s popisem a zpřístupněním záznamů i digitalizovaných rukopisů a jiných historických dokumentů ocení především badatelé, kterým se otvírají nové možnosti pro jejich práci. Nejenže jim je předkládán nástroj pro vyhledání informačních pramenů, často mají ale také možnost v databázi zjistit vazby vyhledaného dokumentu na jiné informační prameny. V neposlední řadě mají možnost podívat se na digitalizovanou kopii dokumentu, je-li dostupná. K originálům je často obtížný přístup. Místo uložení může být prostorově velmi vzdáleno. Digitalizovaná forma dokumentu také není náchylná k poškození při běžné manipulaci a studiu jako originál. A v neposlední řadě může digitalizovaný dokument studovat více badatelů najednou.

6 HLAVIČKY EAD A JEJICH APLIKACE

6.1 Formát EAD

Formát Kódování pro popis archiválií, známý jako EAD (Encoding Archival Description, <http://www.loc.gov/ead/>), vznikl v roce 1993 na Univerzitě v Berkeley původně jako „Berkeley Finding Aid Project“ (BFAP). Jeho cílem bylo umožnit tvorbu registrů, rejstříků, seznamů a dalších pomůcek pro snadnější vyhledávání zdrojů v knihovnách, exponátů muzeí, sbírek rukopisů apod. Pro jednodušší dostupnost materiálů bylo nutné zpřístupnit především informace o vlastnictví (EAD, 2006).

Knihy může být vydána v určitém nákladu. Jednotlivé exempláře jsou v podstatě identické. Na rozdíl od knih jsou archivní exponáty obvykle jedinečné. I když vznikne replika, kopie či jiná reprodukce exempláře, nemá už stejnou hodnotu jako originál.

Dalším specifickým rysem muzejních fondů je jejich účel. Knihovny slouží obvykle především čtenářům a uživatelům. Archivní fondy jsou určeny nejen uživatelům a návštěvníkům, ale především k archivaci lidské kultury, historie a lidského vývoje. Zde uložené materiály bývají vzácné. Některé archiválie je nutné chápat v historickém kontextu jejich vzniku, jinak mohou být špatně vyloženy nebo dokonce zcela nepochopeny (Pitti, 1999).

Bylo stanoveno několik kritérií, která měl formát splňovat:

- schopnost vyjadřovat vzájemné hierarchické vztahy
- schopnost zachovat vztahy v různých stupních popisu
- schopnost znázornit hierarchické vztahy, které získává jedna hierarchická úroveň od jiné úrovně
- možnost přesunů v rámci hierarchických informačních struktur
- podpora charakteristiky prvků při indexaci a vyhledávání
- možnost sdílení stejných dat více vyhledávacími stroji

Volilo se mezi několika formáty - jednoduchý ASCII text, ASCII užívající formátovací značky HTML, formát MARC a SGML. Po analýze byl zvolen SGML, protože nejlépe splňoval funkční požadavky. Vyjadřuje logickou strukturu zdroje

a umožňuje v této struktuře vyhledávat. Jeho značky mohou být vnořeny přímo do zdroje. Naznačuje vztahy mezi částmi zdroje i externí vazby.

V březnu 1995 byla zveřejněna první verze *Berkeley Finding Aid Project DTD* (BFAP). Po testování této verze bylo přistoupeno k revizi a úpravám DTD, datového modelu a souvisejících zdrojů. Skupina odborníků se nechala inspirovat příkladem metadatového formátu TEI a shodla se na tom, že pro vyhledávání jsou důležité informace o zdroji (např. autor, název, datum vytvoření) i údaje o celé skupině zdrojů (archivu, sbírce apod.).

Tato změna struktury ovlivnila stávající prvky modelu BFAP. Ty, které nebyly vypuštěny, byly rozčleněny do dvou skupin - na popisné prvky a všeobecné prvky. Popisné prvky se váží k určitým částem zdroje. Všeobecné prvky se mohou objevit kdekoli ve zdroji. Bylo zároveň rozhodnuto, že pokud se v BFAP objeví velmi podobné údaje, které jsou obsažené ve formátu TEI, budou převzaty ze směrnic TEI.

Model BFAP využil možnosti SGML definovat vlastnosti. Přiřazením vlastnosti k určitému prvku lze zúžit význam daného prvku. Takto je možné vyhnout se definování velkého množství přesně určených podrobných prvků a vytvořit pouze základní soubor obecných prvků s možností blíže je specifikovat použitými vlastnostmi.

Mezi základní změny, které značně zlepšily stávající model BFAP, patří:

- oddělení informací o zdroji a jejich uložení do hlavičky
- rozdělení hierarchicky členěného popisu o jednotce a pomocných informací
- nahrazení modelu BFAP otevřenější strukturou popisných prvků a prvků užitých pro definování zobrazení zdroje

Pro upravený model BFAP se začalo používat označení EAD (Encoded Archival Description). V červenci 1995 se také započalo s přípravami uživatelských manuálů, průvodců a nápověd.

V srpnu 1995 byla na pravidelném setkání Společnosti amerických archivářů (Society of American Archivists, SAA) rozšířena původní pracovní skupina (tzv. Bentley tým) o několik dalších odborníků z různých institucí. Tato nová skupina vznikla pod záštitou Komise pro výměnu archivních informací (Committee on Archival Information Exchange, CAIE), která spadá pod Společnost amerických archivářů. K vývoji se připojila také Kongresová knihovna (Library of Congress),

jejíž kancelář pro vývoj sítí a standardu MARC začala sloužit jako agentura a referenční středisko EAD.

Následovala řada setkání a jednání. Pracovalo se na pravidlech pro popis EAD a průvodcích pro uživatele. Tvořila se knihovna značek EAD. Znovu se revidovalo DTD EAD. Na začátku roku 1996 byl hotový prototyp EAD DTD. Na základě připomínek knihoven a jiných institucí, které testovaly alfa verzi EAD byla připravena již v průběhu roku 1996 beta verze. Po zveřejnění této verze pořádali členové pracovní skupiny četná školení a přednášky. Opět proběhla řada testů a použití EAD bylo ověřeno v praxi. V roce 1998 byla připravena verze 1.0. EAD DTD (EAD, 2006).

V prosinci 2002 byla zveřejněna verze EAD 2002. S touto verzí EAD souvisí některé změny existujících prvků, dále byly zahrnuty nové prvky, vytvořeny nástroje pro tvorbu záznamu v EAD. Pro transformaci záznamů, které již využívají formát XML je možné použít technologii XSLT a převést je tak do XML, HTML, případně jiného formátu. EAD 2002 se stále snaží zachovávat co největší univerzálnost s možností mezinárodního využití a výměny záznamů. EAD je datová struktura a nikoliv standard pro datový obsah. Způsob formulace výrazů v jednotlivých prvcích musí specifikovat národní a mezinárodní standardy. EAD stanovuje pouze typ prvků. Ve verzi EAD 2002 je možné použít prostředí SGML i XML.

6.1.1 Popisné jednotky

EAD nespecifikuje, co má být popisnou jednotkou. Úroveň popisné jednotky stanovují jednotlivé instituce, příp. katalogizátoři, před zahájením zpracování. EAD poskytuje nástroje pro prezentaci hierarchické struktury archivních sbírek a materiálů. Záznamy tak mohou být uspořádány do hierarchických úrovní a mohou mezi nimi být naznačeny vazby.

6.1.2 Struktura formátu EAD

EAD je soubor pravidel pro popis intelektuálních a fyzických částí zdrojů tak, aby byly zdroje pomocí tohoto popisu nalezitelné, jednoznačně určené a zobrazené. Cílem je zpřístupnit uživateli archivní zdroje mnoha institucí. Proto musí být prvky EAD co nejvíce univerzální. EAD slouží především pro popis veřejně dostupných zdrojů (EAD, 2002a). EAD DTD umísťuje obsah položek pro vyhledávání

do jednotlivých prvků. Každý prvek má své jméno, definici a obecnou charakteristiku. Pouze několik prvků je povinných. Ostatní jsou volitelné. Mezi prvky je možné vyjádřit určité vztahy. Prvkům jsou přiřazeny určité vlastnosti, které zužují jejich význam (EAD, 1998). Tabulka s abecedním výčtem prvků EAD je v příloze 9.

EAD 2002 popisuje 146 prvků. Každý prvek má přiřazeno jméno tagu (zkrácené jméno ve strojem čitelné podobě), jméno prvku (plné jméno prvku vystihující jeho podstatu), charakteristiku (základní popis prvku, způsob jeho užití apod.), typ údajů, které může prvek obsahovat (prázdný prvek, text, jiné prvky), vlastnosti (vyjmenování všech vlastností, které mohou být spojeny s prvkem) (EAD, 2002e).

Vlastnosti, které se váží k prvkům, mohou být volitelné, povinné nebo definované v DTD. Většina vlastností je volitelná. Označují se jako #IMPLIED. Vyžadované vlastnosti jsou značeny #REQUIRED. Musí být specifikovány v počátečním tagu prvku. Vlastnosti označené #FIXED mají pouze hodnoty specifikované v DTD.

Každá vlastnost obsahuje několik typů dat. Ta jsou odlišena proto, aby byl obsah snadněji identifikován a zpracován analyzátozem dat. Jde o CDATA, ENTITY, ID, IDREF, IDREFS, NMTOKEN a NMTOKENS. S každým typem dat se nakládá jiným způsobem. U typu CDATA není nutná žádná specifikace. Typ ENTITY se považuje za kódovanou část zdroje a vlastnost s tímto typem dat vyžaduje platné vyjádření. Typ ID je jednoznačně rozpoznán jako identifikátor. Díky tomu může analyzátor provést kontrolu na jeho jedinečnost. IDREF odkazuje na platné identifikátory jiných vlastností. IDREFS obsahuje seznam hodnot odkazů identifikátorů. NMTOKEN je jednoslovná zkratka jména prvku. Typ NMTOKENS obsahuje seznam těchto zkratk (EAD, 1998).

Může být několik typů vlastností prvků - základní, vztahové nebo formátovací. Základní znaky slouží např. k vyhledávání, kontrole přístupových práv, bibliografickým odkazům na jiný zdroj nebo jeho část, určení kódování, jazyka zdroje, použitých pravidel popisu a schémat (AACR2, MESH, SCSH,...) nebo úrovně popisu (EAD, 1998). Vztahové znaky se používají pro odkazy a vyjádření s nimi spojených údajů, např. požadovaný software, identifikátory, typ odkazu (na část užívaného zdroje, na externí zdroj), názvy vzdálených zdrojů, apod.

(EAD, 1998). Formátovací znaky ovlivňují strukturu a vzhled celého zdroje (zarovnávání textu, typy písma, odstavce,...) (EAD, 1998).

Obdobně jako formát TEI slouží i formát EAD k popisu celého zdroje. Struktura záznamu EAD je rozčleněna na dvě části. První částí je popis jednotky a druhou částí jsou údaje o archivu, fondu či jiném souboru zdrojů. První část obsahuje dva nejvyšší prvky - hlavičku EAD Header <eadheader> a prvek Front Matter <frontmatter>. Druhá část obsahuje třetí nejvyšší prvek Archival Description <archdesc>. Tyto tři prvky jsou vnořeny do celkového popisu ohraničeného tagem Encoded Archival Description <ead>.

Záznamy jsou uspořádány do hierarchických úrovní a jsou mezi nimi naznačeny vztahy. Z celku vedou odkazy na jednotlivé části. Odkazy mohou být cíleny v rámci jednoho archivu nebo na externí objekty. Existují i speciální tagy pro zajištění propojení s autoritami (EAD, 1998). Tagy pro odkazy jsou označeny značkou <ref>. Užívají se pro odkazy mezi sekcemi záznamu i mezi různými záznamy.

Přestože se formát EAD zaměřuje na obsah zdroje a nikoliv na jeho vzhled, počítá se s použitím stylů při zobrazování textů. Styly hrají také důležitou roli při transformacích textů kódovaných pomocí EAD v XML do jiných formátů (např. HTML).

6.1.3 Hlavičky EAD

Hlavička EAD vychází z hlavičky TEI. Zachycuje informace důležité pro uložení a následné vyhledávání zdroje. Jsou to informace o vzniku, úpravách, vydání a přístupu ke zdroji. Sestává ze čtyř částí, které mohou být dále členěny:

- Identifikátor EAD <eadid> - jednoznačný identifikátor zdroje
- Popis souboru <filedesc> - bibliografické informace o zdroji členěné do hierarchicky dělených prvků <titlestmt> (hlavní název, podnázev a autor) a <publicationstmt> (údaje o nakladateli, datum a místo vydání)
- Profil zdroje <profiledesc> - původce kódované verze zdroje, důvod kódování, jazyk apod. s prvky <creation> (individuální odpovědnost za vytvoření díla spolu s údajem o datu) a <language> (jazyk záznamu)
- Historie změn <revisiondesc> - shrnutí všech podstatných úprav zdroje

Uvedené části by za sebou měly následovat v pořadí, které udává DTD EAD. EAD proto dovoluje užít volitelný prvek <frontmatter> pro různé další informace uvedené na titulní straně a v úvodních částech zdroje. Jde např. o úvod, věnování, návod k použití a další texty spojené se vznikem, vydáním či používáním zdroje.

V prvku <archdesc>, který v popisu stojí na stejné úrovni jako hlavička TEI, se nacházejí strukturované popisné informace. Informace jsou uspořádány v několika hierarchických úrovních. Mezi nejdůležitější prvky vyšší úrovně patří prvek <did> (Descriptive identification). Slouží k identifikaci zdroje a rozhodnutí uživatele o relevanci zdroje vzhledem k jeho potřebám. Hierarchicky nižšími prvky spadajícími pod prvek <did> jsou např. název, autor, datum vytvoření, údaje o vlastnictví zdroje,... (Pitti, 1999). Z vyšší úrovně popisu na nižší úroveň jsou „děděny“ informace. Je zde množství volitelných prvků pro bližší popis jednotlivých částí celku i pro administrativní informace typu akvizičních údajů či přístupových práv. Pro sdružení prvků o jedné jednotce se používá ohraničující tag Popis podřízených složek <dsc> (Description of Subordinate Components) (EAD, 1998).

Hlavička v příkladu 16 obsahuje první tři části hlavičky EAD zmíněné výše (EAD, 2002c). Všechny části jsou dále strukturovány, obsahují dceřinné prvky. Poslední část <revisiondesc> není v příkladu uvedena.

Příklad 16: Hlavička EAD

```
<eadheader audience="internal" countryencoding="iso3166-1" dateencoding="iso8601"
langencoding="iso639-2b" repositoryencoding="iso15511">
  <eadid countrycode="us" mainagencycode="cu-i" publicid="//us::cu-i//TEXT us::cu-i:: p29.sgm//EN">Mildred Davenport
  Dance Programs and Dance School Materials, MS-P29
</eadid>
<filedesc>
  <titlestmt>
    <titleproper>Guide to the Mildred Davenport Dance Programs and Dance SchoolMaterials</titleproper>
    <author>Processed by Adrian Turner; machine-readable finding aid created by AdrianTurner</author>
  </titlestmt>
  <publicationstmt>&hdr-cu-i-spcoll;
    <date>&copy; 2001</date>
    <p>The Regents of the University of California. All rights reserved.</p>
  </publicationstmt>
  <notestmt>
    <note>
      <p>
        <subject source="cdl">Arts and Humanities--Dance--Dance Performance</subject>
        <subject source="cdl">Arts and Humanities--Dance--Dance History andCriticism</subject>
        <subject source="cdl">Area, Interdisciplinary, and Ethnic Studies--AfricanAmerican Studies</subject>
      </p>
    </note>
  </notestmt>
</filedesc>
<profiledesc>
  <creation>Machine-readable finding aid derived from MS Word. Date of source:
  <date>2001.</date></creation>
  <language>Description is in <language>English.</language>
</profiledesc>
</eadheader>
```

6.1.4 Shrnutí informací k formátu EAD

Formát EAD i jeho struktura jsou složité. Prvky v záznamech mohou být uloženy v několika hierarchických úrovních a některé prvky mohou být uloženy na různých místech jako podřazené prvky více částí popisu. V popisu je možné používat také prvky s odkazy nejen na různé části popisovaného materiálu ale také na externí zdroje. Dalším rysem tohoto formátu je přítomnost prvků, které určují formátování a vzhled materiálu. Používají se především pro tabulkové zobrazení textu.

Elektronický popis muzejních materiálů má kratší historii než elektronický popis v knihovnictví. Přesto se v současnosti rychle rozvíjí a vzniká řada různých projektů a systémů zaměřených na tuto oblast. Archiváři si uvědomují význam popisu archiválií a jeho zpřístupňování pro uchování a snadnější dostupnost sbírek.

6.2 Centrála archivů



Centrála archivů (Archives Hub,

<http://www.archiveshub.ac.uk/>) je webová služba, která nabízí volný přístup k popisným informacím o archivech řady britských vysokoškolských institucí z oblasti humanitních a filozofických věd (Ramsden, 2003). Jde většinou o popisy na úrovni sbírky (Archives Hub, 2000d). Tyto sbírky jsou výsledkem činnosti určité osoby nebo organizace. Mohou obsahovat listiny, knihy, fotografie, audiovizuální nahrávky i jiné typy materiálů (Archives Hub, 2000d). V případě, že jednotlivé instituce zpřístupňují také záznamy na nižších úrovních (popis jednotky), jsou dostupné zprostředkovaně i tyto záznamy (Archives Hub, 2000d). K vybraným zdrojům musí uživatelé přistupovat přes jednotlivé archivy, jejichž seznam s odkazy je zpřístupněn na stránkách Centrály archivů (Archives Hub, 2000d).

Kromě informací o archivech poskytuje Centrála archivů také informace a školení katalogizace archivních sbírek s použitím XML a metadatového formátu EAD.

Centrála archivů je spolu s dalšími archivy součástí národní sítě archivů Velké Británie. Vývoj Centrály archivů je řízen komisí složenou ze zástupců spolupracujících institucí, Národním archivem a zástupci dalších archivů (Archives Hub, 2000d). Do vývoje této služby zasahují pomocí veřejných diskusí také uživatelé a různí přispěvatelé z řad veřejnosti (Archives Hub, 2000d). Ukázalo se, že připomínky spolupracujících organizací jsou pro provoz a vývoj systému velmi důležité (Hill, 2002).

Systém Centrála archivů je provozován na serverech MIMAS²⁶ na Univerzitě v Manchesteru pod záštitou Konsorcia univerzitních vědeckých knihoven (Consortium of University Research Libraries, CURL). Hlavním spozorem je Spojená komise informačních systémů (Joint Information Systems Committee, JISC), organizace, která podporuje využívání informačních systémů a technologií

²⁶ Informační a přidružený servis v Manchesteru MIMAS (Manchester Information & Associated Services) je národní datové centrum, které zajišťuje síťový přístup ke klíčovým informačním zdrojům pro podporu vysokoškolského vzdělávání ve Velké Británii (MIMAS, 2002).

v oblasti vzdělávání ve Velké Británii. Vývoj systému zajišťuje Univerzita v Liverpoolu (Archives Hub, 2002b).

Pilotní fáze projektu Centrála archivů probíhala od září 1999 do července 2000. Šlo především o průzkumy v oblasti akademické a archivační. Pilotní fáze se účastnilo patnáct vysokoškolských knihoven a archivů. Výsledkem bylo zhruba 3 000 záznamů na úrovni archivů a prototyp systému. Na základě výsledků pilotního projektu začala od roku 2001 Centrála archivů fungovat jako webová služba s finanční podporou JISC. Původně se projektu účastnily pouze vysokoškolské instituce a instituce postgraduálního vzdělávání. Nyní je však zapojeno přes 150 vzdělávacích institucí různé velikosti a charakteru.

Popisné záznamy jsou vytvářeny v jednotlivých institucích, které sbírkou vlastní, a jsou zasílány do Centrály archivů. Pro vytváření záznamů se používají formuláře, které jsou strukturované podle normy Mezinárodní standard pro popis archiválií ISAD(G) (General International Standard of Archival Description). ISAD(G) klade důraz na víceúrovňový popis a vztahy mezi bibliografickými záznamy a autoritními záznamy, které jsou vytvářeny dle ISAAR(CPF)²⁷ (ISAD(G), 2000). Použit je metadatový formát EAD uložený ve standardu XML.

Aby se předešlo v co největší míře nejednotnosti a různému přístupu jednotlivých katalogizátorů, jsou určitá pole povinná. Dále mají katalogizátoři k dispozici informace, jak nejlépe záznam připravit. Před přidáním do databáze je každý záznam zkontrolován editorem.

Největší rozdíly lze najít v oblasti věcného zpracování, protože různé instituce využívají odlišné tezaury a hesláře. Kromě předmětových hesel Kongresové knihovny jsou používány např. tezaurus UNESCO nebo z něho odvozený tezaurus UKAT (United Kingdom Archival Tezaurus). Nejednotnost v oblasti věcného popisu se snaží řešit tzv. „Subject Finder“, který používá metodu shlukování (clustering). Aplikuje ji na použité výrazy z tezaurů a výrazy z volných polí s věcným popisem. Odpadá tak nutnost vyhledávání podle více řízených slovníků (Hill, 2006).

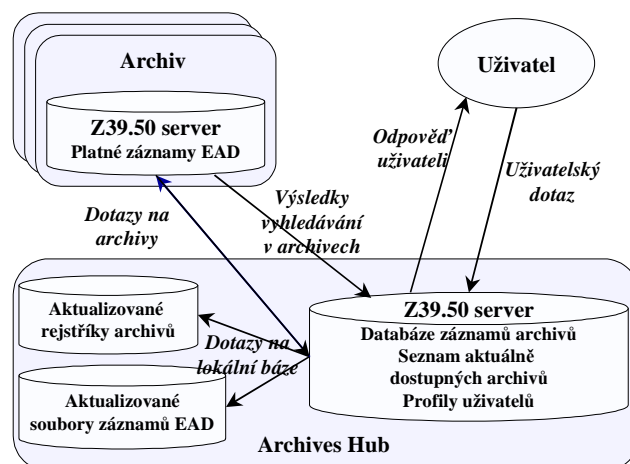
Využití Centrály archivů od roku 2001 stále vzrůstá. Výrazný růst byl zaznamenán poté, co začaly být koncem roku 2002 zpřístupňovány popisné záznamy přes webové vyhledávače.

²⁷ Mezinárodní standard pro popis archivních autoritních záznamů - korporace, osobní a rodová jména (International Standard Archival Authority Record for Corporate Bodies, Persons and Families, ISAAR(CPF)) - řeší otázky způsobu zápisu jmen korporací a osob, včetně odkazů typu viz a viz též.

Do července 2005 byla databáze Centrály archivů spravována plně centrálně na serverech MIMAS na Univerzitě v Manchesteru. Změnu ale přinesl software Spokes, který dovoluje každé instituci spravovat si lokálně svůj vlastní archiv a pouze ho Centrále zpřístupňovat přes protokol Z39.50 (Hill, 2005). Spokes zahrnuje relační databázi, která obsahuje informace o každém vzdáleném archivu spolu s údaji, kdy byl tento archiv naposledy navštíven, aktualizován, zda je aktuálně dostupný atd. V případě delší nedostupnosti lokálního archivu je možné přesměrovat odkazy na kopii archivu na centrálním serveru (Hill, 2005).

Schéma tohoto distribuovaného systému, který zahrnuje kromě centrální databáze lokální archivy a dovoluje v nich souběžně vyhledávat a vyměňovat a sdílet mezi nimi záznamy, je uvedené na obrázku 9.

Obr. 9: Schéma distribuovaného systému Centrály archivů



Za současné spolupracující organizace, které přispívají svými záznamy archivů do systému, jmenujme z rozsáhlého seznamu alespoň vybrané instituce a zaměření jejich archivů (Archives Hub, 2000a).

- Univerzita v Aberdeenu - historické a kulturní archivy týkající se severovýchodního Skotska
- Univerzita ve Walesu - akademické práce a archivy, které vytvořili významné osobnosti
- Univerzita v Bathu - vědecké archivy
- Univerzita v Birminghamu - přes 300 archivů, mezi něž se řadí např. Chamberlainův archiv, archiv organizace YMCA, univerzitní archiv, archiv náboženských textů nebo archiv divadelních studií

- Oxfordská Bodleiana - literární, historické, politické archivy z 19. a 20. století
- Univerzita v Glasgow - archiv divadelních textů a archiv s tématy lékařství
- Umělecká škola v Glasgow - kolem 300 archivů k oblasti umění a architektury

Centrála archivů představuje zdarma dostupné rozhraní pro přístup k mnoha různým archivům. Tematicky pokrývá Centrála archivů archivy vědecké, politické, literární, historické, archivy průmyslového rozvoje, medicíny, historie diplomacie, umění a architektury. Archivy, které Centrála archivů popisuje, obsahují nejen listiny a podobné materiály, ale také např. filmy a audio nahrávky, obrazy a další objekty výtvarného umění a dokonce rentgenové snímky (Archives Hub, 2002b).

Popis archivů provádějí informační pracovníci. Popis zahrnuje informace o celém archivu (způsob organizace materiálů), informace o typech záznamů v archivu, předmětovém a časovém pokrytí objektů, informace o možnostech a omezeních přístupu do archivu a předmětová hesla (osoby, místa, tematické vymezení) (Archives Hub, 2002b).

6.2.1 Popisné jednotky

Popis je prováděn na úrovni archivů ve smyslu sbírek archiválií. Archivy mohou být uspořádány dle potřeb každé instituce (místně, regionálně, dle formátu, tematicky, dle autora, ...). Důležité však je, že popisované archivy musí být dostupné externím uživatelům. Prozatím jde z velké části o fyzické archivy a sbírky, které jsou dle možností jednotlivých institucí digitalizovány.

Pokud některá instituce vytváří záznamy na nižší úrovni (úrovni jednotky), Centrála tyto záznamy zprostředkovaně zpřístupní také.

6.2.2 Metadatové záznamy

Záznamy jsou tvořeny pomocí webového formuláře. Do formulářů se podle pravidel ISAD(G) vyplňují údaje, které jsou automaticky převáděny do XML souborů, v nichž jsou údaje uloženy v metadatovém formátu EAD. K jednotlivým polím je možné přes hypertext vyvolat kontextovou nápovědu. Pro generování stránek s příslušnými prvky EAD jsou využívány JavaScripty.

V úvodu formuláře katalogizátor volí úroveň popisu. Je možné vytvářet popis pouze na úrovni archivu či sbírky, dále je možné kombinovat tuto úroveň s popisem na nižších úrovních, nebo je možné zvolit popis pouze na nižší úrovni. I v případě

vytváření záznamu na nejnižší úrovni je potřeba v záznamu uvést základní údaje o archivu či sbírce, do níž jednotka patří. Je to proto, aby byl záznam jasný a srozumitelný i v případě, že bude zobrazen samostatně bez souhrnného záznamu archivu (Archives Hub, 2000c). Poslední část formuláře tvoří prvky pro věcný popis (jména autorů a korporací), které by měly být vybírány z tezaurů a řízených slovníků. V tabulce 8 jsou uvedeny prvky, které je možné do webového formuláře vyplnit (Archives Hub, 2000c).

Tab. 8: Prvky formátu používaného Centrálou archivů dle webového formuláře

Pole formuláře Centrály archivů	Prvky EAD	Poznámky
Jméno repozitáře	<repository>	
Identifikační údaje		
Referenční kód	<eadid>	kód země, číslo archivu, kód sbírky, např. GB 0033 AHM ²⁸
Název	<titlemt>	co nejvýstižnější název archivu, sbírky nebo jednotky
Datum vytvoření	<unitdate>	není-li znám přesný datum, doporučuje se uvést přibližné datum či století, forma vyjádření je volná
Normalizované datum	<unitdate normal="">	normalizované datum, např. <unitdate normal="1201-1299">13th century</unitdate>
Stupeň popisu	<archdesc level="fonds">	automaticky generované pole
Rozsah popisné jednotky	<extent>	vyjádření velikosti zdroje i s jednotkami velikosti či rozsahu
Kontextové údaje		
Jméno tvůrce	<origination>	jméno autora, autoritní forma jména je uvedena v prvku <persname>
Administrativní/biografická historická poznámka	<bioghist>	poznámka ke kontextu objektu
Archivní historie	<custodhist>	
Přímý zdroj akvizice	<acqinfo>	
Údaje k obsahu a struktuře		
Účel a obsah	<scopecontent>	krátká poznámka k obsahu
Ocenění, zhodnocení	<appraisal>	
Růst hodnoty díla	<accruals>	
Způsob uspořádání	<arrangement>	
Podmínky přístupu a využití		
Podmínky přístupu	<accessrestrict>	
Podmínky reprodukce	<userestrict>	
Kód jazyku zdroje	<langmaterial>	trojznakový kód či více kódů dle ISO 639-2
Fyzické znaky	<phystech>	
Vyhledávací nástroje	<otherfindaid>	např. poznámka o existenci lístkového katalogu
Údaje o souvisejících zdrojích a materiálech		
Údaje o existenci a lokaci originálu	<originalsloc>	
Údaje o existenci a lokaci kopie	<altformavail>	
Příbuzná jednotka	<relatedmaterial>	
Dílo pojednávající o nebo zakládající se napopisovaném zdroji	<bibliography>	
Poznámky		
Poznámka	<note>	
Popisné údaje		
Poznámka archiváře	<processinfo>	autor záznamu, zdroje informací v záznamu
Datum vytvoření záznamu	<date>	pole generováno automaticky
Přístupové body osobní jméno jméno rodu jméno korporace jméno místa, oblasti předmět	<controlaccess> <persname> <famname> <corpname> <geogname> <subject>	Prvek propojuje všechny záznamy jednoho autora, o jednom tématu apod. Jeho obsah musí být kontrolován řízenými slovníky, tezauri či soubory autorit. Obsahuje údaje z prvků <unittitle>, <originator>, <bioghist>.

²⁸ Referenční číslo odkazuje na kontaktní informace na archiv nebo knihovnu.

Údaje, které byly zapsány do formuláře, jsou převedeny do formátu EAD a takto vytvořený záznam se otevře v novém okně. Záznam vytvořený pomocí webového formuláře je možné editovat pomocí jakéhokoliv textového nebo XML editoru (Archives Hub, 2000f). Kromě úpravy obsahu prvků, přidávání nebo mazání prvků je možné ke každému prvku připojit atribut <audience>, který dokáže zajistit skrytí údajů, které nemají být veřejně přístupné, např. <profiledesc audience="internal">.

Centrála archivů definuje podle pravidel ISAD(G) povinná pole, která jsou uvedena v tabulce 9 spolu s pravidly pro formátování obsahu těchto polí (Collis, 2001).

Tab. 9: Pole ISAD(G), které povinně obsahují záznamy Centrála archivů

ISAD(G)	formulář Centrály archivů	Kontrolovaný formát dat
3.1.1	Referenční kód (reference code)	ISO 3166 ARCHON
3.1.2	Název (title)	-
3.1.3	Data vytvoření (dates of creation)	AACR2, pravidla NCA
	Normalizované datum (normalised date)	ISO 8601
3.1.4	Stupeň popisu (level of description)	-
3.1.5	Rozsah popisné jednotky (extent of unit description)	-
3.2.1	Jméno tvůrce (name of creator)	lokální formátování
3.2.2	Administrativní/biografická historie (administrative/biographical history)	-
3.3.1	Účel a obsah (scope and content)	-
3.4.1	Podmínky přístupu (conditions governing access)	-
3.4.3	Jazyk materiálu (language of material)	ISO 639-2
3.4.5	Vyhledávací nástroje (finding aids)	-
3.7.1	Poznámka archiváře (archivist's note)	-
3.7.3	Datum vytvoření záznamu (date(s) of description)	-
-	Přístupové body (access points) Minimální požadavky: uvést jméno autora ve formě podle řízených slovníků [ISAAR(CPF)]	Soubory jmenných autorit NRA tezaurus Unesco UKAT LCSH AAT TGM II

Příloha 10 obsahuje záznam archivu ve formátu EAD vygenerovaný pomocí formuláře Centrály archivů podle pravidel ISAD(G) (Archives Hub, 2000e). Kromě běžných údajů (název, technické údaje, datum posledních úprav apod.) dokládá příklad obsáhlé biografické a administrativní poznámky ke kontextu objektu (prvek <bioghist>) a dále poznámky k účelu a obsahu archivu (prvek <scopecontent>). Údaje v těchto prvcích jsou děleny do několika odstavců. Na konci záznamu je patrné použití prvku <emph>, který upřesňuje a více strukturuje údaje v prvcích <geogname>, <subject>, <persname> a <corpname> (spadají pod prvek <controlaccess>).

Centrála archivů umožňuje popis archivu na více úrovních. Jde o případy, kdy archiv obsahuje menší ucelené sbírky. Pro popis na více úrovních popisu je možné užít nejdříve formulář pro nejvyšší úroveň popisu a následně pro každý podřazený záznam formulář pro nižší úroveň popisu. Soubory, které takto vzniknou, vytvoří komplexní platný dokument EAD (Collis, 2001).

Formát EAD byl pro účely systému Centrály archivů pozměněn zcela nepatrně. Cílem těchto změn bylo zajištění větší konzistence popisu. Mezi úpravy patří změna prvku <controlaccess>, do jehož podřízených prvků <geogname>, <subject> a <persname> byl přidán prvek <emph> (emphasis) pro užší vymezení významu částí těchto prvků. Použití tohoto prvku formát EAD 2002 povoluje (Hill, 2002).

Centrála archivů neomezuje spolupracující instituce ve využití potřebných tezaurů, předmětových hesel a rejstříků, jako jsou např. tezaurus Unesco²⁹, tezaurus UKAT³⁰, tezaurus geografických jmen TGN (Thesaurus of Geographic Names)³¹, předmětová hesla Kongresové knihovny (LCSH). Použít je možné i jiné tezaury (AAT³², ERIC³³, HASSET³⁴, MeSH³⁵, ...) Pro jmenné autority využívá Centrála archivů jmenné soubory autorit Kongresové knihovny nebo referenční kódy NRA. Zpracování se řídí pravidly AACR2 a Pravidly archivního popisu NCA Národní rady archivů (National Council of Archives) (Archives Hub, 2000b).

²⁹ Tezaurus UNESCO je řízený slovník vyvíjený Organizací spojených národů pro vzdělávání vědu a kulturu (United Nations Educational, Scientific and Cultural Organisation). Obsahuje termíny z oblasti výchovy a vzdělávání, vědy, kultury, sociálních a humanitních věd, informační vědy a komunikací, politiky, práva a ekonomie. Tezaurus obsahuje ekvivalenty anglických termínů ve francouzštině a španělštině (UNESCO, 2005).

³⁰ UK Archival Tezaurus byl vytvořený pro sektor archivů ve Velké Británii.

³¹ TGN je tezaurus obsahující přes milion jmen a dalších informací o geografických místech. Obsahuje jména míst současného světa i historická jména. Zahrnuje také jména měst a národností. Termíny jsou specifikovány jednoznačnými identifikátory a místa jsou rozlišena typy (About the TGN, 2006).

³² Art & Architecture Thesaurus – tezaurus pro umění a architekturu.

³³ US Educational Resources Information Center (ERIC) Thesaurus – tezaurus pro oblast vzdělávání a výchovy.

³⁴ Humanities and Social Sciences Electronic Thesaurus – tezaurus z oblasti humanitních a sociálních věd.

³⁵ Medical Subject Headings – předmětový heslář pro lékařství.

6.2.3 Vyhledávání

Pro vyhledávání používá Centrála archivů vyhledávací systém Cheshire, který vyvinul profesor Ray Larson z Kalifornské univerzity v Berkeley. Na jeho dalších úpravách a vývoji klienta se podíleli pracovníci Univerzity v Liverpoolu. Systém byl od počátku určen k použití s formátem EAD.



Jde o robota, který indexuje jednotlivé XML soubory s metadatovými záznamy v EAD. Tyto záznamy archivů dodávají jednotlivé instituce, pod něž dané archivy a sbírky spadají. Z každého záznamu jsou vytvořeny podle obsažených prvků rejstříky, ve kterých je následně možné vyhledávat. Vyhledávání se neomezuje pouze na rejstříky, ale je možné vyhledávat i v režimu volného textu.

Do července 2005 byla databáze Centrály archivů pouze jedna centrální. Pak ale systém přešel na distribuovaný model, který dovoluje institucím spravovat vlastní databáze a repozitáře a data z nich pouze zprostředkovávat přes protokol Z39.50. Spolupráci s centrálním vyhledávacím systémem zajišťuje software Spokes. Každou noc probíhá synchronizace a aktualizace rejstříků z jednotlivých institucí spravovaných softwarem Spokes s centrálními rejstříky.

V současnosti je používána verze 3 systému Cheshire³⁶, která spolu se systémem Spokes plně podporuje formát XML. Systém se vyvíjí tak, aby bylo možné záznamy nejen vyhledávat a zobrazovat, ale také stahovat přes Z39.50 a OAI-PMH ve formátu EAD i dalších rozšířených formátech typu Dublin Core nebo MARC. Zajištěna by měla být také podpora standardu METS pro přenos záznamů (Archives Hub, 2000g).

Systém Centrály archivů nabízí jednoduché rozhraní pro vyhledávání a pokročilé vyhledávání. Pokročilé vyhledávání umožňuje vyhledávat ve vybraných polích nebo v plném textu a užívat booleovské operátory (and/or). Důležitá je možnost omezení vyhledávání na vybraný archiv. Známe-li referenční číslo zdroje, lze vyhledávat i podle něho.

Další možností je listování v rejstřících názvů sbírek a archivů, předmětů, jmen osobních a korporativních, geografických jmen, osobních přízvisek a referenčních čísel.

³⁶ Více informací o systému Cheshire 3 na adrese <http://cheshire3.sourceforge.net/>.

Zvláštní pomůckou vyhledávání je nástroj „předmětové vyhledávání“ (subject finding). Jak již bylo uvedeno výše, používá metodu shlukování. Vyhledává k zadanému výrazu související výrazy v existujících rejstřících a tezaurech. Systém nerozlišuje velká a malá písmena. Automaticky vyhledává jiné tvary slova (nurse, nursing, nurses apod.) (Archives Hub, 2002a).

Obr. 10: Záznam s více úrovněmi popisu

<div data-bbox="358 516 493 621"> </div> <div data-bbox="542 548 719 569"> <p>Table of Contents</p> </div> <div data-bbox="380 627 760 1045"> <ul style="list-style-type: none"> University of Edinburgh + University Governance and legal basis of the university + Administration and Support Services + Academic Administration - Academic Units <ul style="list-style-type: none"> Institute of Animal Genetics - Institute of Cell and Molecular Biology <ul style="list-style-type: none"> 'SPACE' Planning Executive Committee School of Biology: Planning and Policy Group and Steering Committee Division of Biological Sciences Working and Advisory Group + Science Studies Unit </div>	<div data-bbox="971 525 1203 546"> <p>University of Edinburgh</p> </div> <div data-bbox="792 590 1344 808"> <p>Held at: Edinburgh University Library, Special Collections Reference and contact details: GB 237 EUA IN1 Title: University of Edinburgh Dates of Creation: 1582-date Extent: (many hundred) linear metres Name of Creator: University of Edinburgh 1583-: university Edinburgh (Scotland) Language of Material: This material is in English Level of Description: fonds Revisions:</p> </div> <div data-bbox="792 831 1271 852"> <p>Published by: Edinburgh University Library, Special Collections</p> </div> <div data-bbox="792 875 974 982"> <p><i>University of Edinburgh George Square Edinburgh EH8 9LJ Scotland United Kingdom</i></p> </div> <div data-bbox="979 1005 1196 1026"> <p>Administrative History</p> </div> <div data-bbox="792 1050 1382 1377"> <p>The University of Edinburgh was established by Royal Charter in 1582 . It was originally called Tounis College, when part of a legacy left by Robert Reid, Bishop of Orkney in 1558 had established a college of which the Town Council had gained control to establish a College of Law on the South side of Edinburgh. The inception of the University took place in 1583 . In 1617 when King James VI of Scotland (I of England) visited the College it was decreed that the College should change its name to King James' College, although the College continued to use the older title. The first change in the corporate body of the University was not until 1935 when the first merger took place. This was between the Faculty of Divinity of the University of Edinburgh and New College. This was due to the re-union of the Church of Scotland in 1932. The next merger was in 1951 when the Royal (Dick) Veterinary School was reconstituted as part of the University of Edinburgh. The Royal (Dick) Veterinary School achieved full faculty status in 1964. In 1998 Moray House Institute of Education became the Faculty of Education.</p> </div> <div data-bbox="792 1398 1382 1591"> <p>The first classes of the university were held in Hamilton House known as the Duke's Lodge. In 1582 a site that included St Mary in the Fields was acquired. Many new buildings and extensions were made to the site of Hamilton House after 1616. Two prominent stages of building for the University were those undertaken by Robert Adam and William Playfair. In 1869 the site next to the Edinburgh Royal Infirmary was acquired. Building on this project was completed by the end of the 19th century . The University today is situated around these areas in the centre of Edinburgh and Kings Buildings and there are also campuses at Holyrood and elsewhere.</p> </div>
--	---

Po vyhledání se zobrazí okno rozdělené na dvě části. V levé části je seznam vyhledaných záznamů, v pravé části okna pak konkrétní vyhledaný záznam. Obvykle jde o záznam celého archivu nebo sbírky. Může jít ale také o záznam s více úrovněmi popisu. Pak je seznam vyhledaných výsledků v levé části okna nahrazen stromem

s jednotlivými částmi popisu. Části stromu je možné rozbalovat a přes odkazy, které jsou v nich umístěné, se posouvat přímo na vybranou úroveň a část popisu v hierarchickém záznamu. Příklad takového záznamu je vidět na obrázku 10.

6.2.4 Shrnutí informací k Centrále archivů

Centrála archivů je součástí širší sítě, která ve Velké Británii eviduje informační zdroje k různým oborům lidské činnosti. Pro oblast humanitních a filozofických věd se stala nezastupitelným informačním zdrojem, počet jejích uživatelů nejen z univerzitního prostředí stále stoupá.

Centrála hraje významnou roli také jako koordinační a vzdělávací centrum a pomáhá ostatním institucím budovat vlastní repozitáře a sbližovat se s problematikou digitálních archivů, metadatových formátů a dalších souvisejících témat.

Pro zájemce je k dispozici také blog k Centrále archivů, mailing list, každý měsíc je vybírána nejzajímavější sbírka měsíce. Je možné zajistit si novinky přes službu RSS.

Díky spolupráci s technicky orientovanými odborníky z Univerzity v Manchesteru udržuje Centrála archivů krok s novými technologiemi, nasazuje nové verze používaného softwaru a zakomponovává do procesu svého fungování i nové systémy. Přejít z centralizované vytvářené databáze na distribuovaně pracující systém je důkazem rozvoje celého systému. Dává více volnosti jednotlivým spolupracujícím institucím.

Také podpora sdílení záznamů přes protokoly Z39.50 a OAI-PMH a podpora standardů jako jsou metadatový formát Dublin Core nebo standard METS naznačují, že Centrála archivů se bude i nadále vyvíjet, sloužit svým uživatelům a spolupracovat s ostatními institucemi na rozvoji informační infrastruktury.

6.3 JANUS



Historie britského projektu Janus (<http://janus.lib.cam.ac.uk/>) sahá do října 2002. Cílem tohoto projektu bylo zajistit jednotný přístup ke katalogům všech archivů a rukopisných sbírek v Cambridge. Do databáze jsou postupně přidávány archivy jednotlivých spolupracujících univerzit a dalších kulturních institucí. Koncem roku 2006 dosáhl jejich počet třicítky a počet zpřístupňovaných katalogů čísla 1500.

Janus je hostován a řízen knihovnou Univerzity v Cambridge. Na financování projektu se podílí řada nejen univerzitních a muzejních institucí. Projekt sponzorují také nadace (Foyle Foundation a Heritage Lottery Fund) (Janus, 2006a).

Projekt nese jméno římského boha bran a dveří. Bůh Janus bývá zobrazován s dvěma tvářemi, z nichž jedna hledí vpřed a druhá vzad, stejně jako dvě strany dveří. Tato symbolika má být obsažena také v projektu samotném. Skrze moderní technologie (pohled vpřed) se uživatelům nabízí možnost nahlédnout do archivů a sbírek nejen historických dokumentů (pohled vzad).

Databáze neobsahuje plné texty dokumentů ani jejich digitální obrazy. Přístup k vybranému dokumentu si musí uživatel zajistit na základě domluvy s institucí, která sbírku či archiv vlastní. Kontaktní informace Janus nabízí.

Většina zúčastněných institucí používá pro katalogizaci aplikaci MS Access. Ostatní instituce vytváří záznamy pomocí formuláře na stránkách projektu, nebo katalogizují své záznamy ručně. Katalogizace probíhá podle ISAD(G) a záznamy jsou uloženy v metadatovém formátu EAD. Pro osobní personální a geografické autority se používají pravidla NCA (National Council on Archives) a pro další věcné zpracování předmětový tezaurus UNESCO a tezaurus geografických výrazů Getty.

Každá spolupracující instituce může na interních stránkách projektu provést upload záznamů do svého adresáře. Z něho jsou uložené záznamy přesouvány do veřejné databáze. Před importem do databáze je ale každý záznam kontrolován. Kontroluje se validita EAD záznamu. Záznamy jsou indexovány, aby byly následně vyhledatelné. Každá instituce může své záznamy upravovat, updatovat nebo i mazat (Janus, 2006a).

Janus používá tzv. „čistič“ („tidyer“). Ten zajišťuje úpravy EAD záznamů. Např. odstraňuje nadbytečné mezery, opravuje DTD, upravuje některé kódování pro XML (např. & amp;) nebo koriguje výrazy z tezaurů a autoritních rejstříků. Dokáže také konvertovat data z jiných archivů (např. z Archives Hub) tak, aby byla vhodná pro Janus (Janus, 2006f).

V první fázi se podařilo vytvořit infrastrukturu, která zajišťuje listování v rejstřících repozitářů a zúčastněných institucí a vyhledávání v metadatových záznamech. V druhé fázi byl pro Janus vytvořen vyhledávací stroj a nástroje, které dovolí připojovat katalogy vznikající v rámci jiných projektů zaměřených na archivy, kterých je ve Velké Británii několik.

6.3.1 Popisné jednotky

V databázi Janus jsou obsaženy především záznamy, které na nejvyšší úrovni popisují archivy a sbírky. U některých archivů neexistují záznamy na nižší úrovni. U většiny jde ale popis na nižší úrovni. Vybraný archiv nebo sbírku je možné samostatně prohledávat a dostat se až k zobrazení záznamu jednotlivého dokumentu. Systém nabízí také vyhledávání napříč různými archivy a sbírkami (Janus, 2006a).

Archivem či sbírkou se rozumí kolekce dokumentů, které shromáždil jednotlivec, rodina nebo instituce. Janus neobsahuje pouze textové archivy a sbírky. Zahrnuje také sbírky audio a video nahrávek, map, fotografií atd. Stejně tak se nezaměřuje pouze na historické sbírky, ale eviduje i sbírky novodobé (Janus, 2006e).

6.3.2 Metadatové záznamy

Jak již bylo uvedeno výše, záznamy jsou vytvářeny podle pravidel ISAD(G) v metadatovém formátu EAD. Tvorba probíhá pomocí nástrojů v MS Access, webového formuláře projektu Janus nebo v textovém, příp. XML editoru (Janus, 2006d). Záznamy jsou uloženy v domovských adresářích jednotlivých institucí a ve veřejné databázi.

Každý metadatový záznam obsahuje hlavičku EAD, která je zásadní pro ověřování záznamu, jeho výměnu, aktualizaci a určení původu. Příklad 17 obsahuje ukázkou takové hlavičky. Jsou v ní uvedeny základní údaje o popisované sbírce či archivu. Obsáhlejší popis může být obsažen v prvku <bioghist>. Záznam

může obsahovat také informace o omezeném přístupu nebo využití zdroje. Přítomny mohou být samozřejmě také různé poznámky. Na základě řízených rejstříků a tezaurů se v záznamu vytváří část <controllaccess>, která odkazuje na autoritní formy jmen autorů nebo geografických jmen (Janus, 2006d).

Příklad 17: Hlavička EAD použitá v databázi Janus (Janus, 2006d)

```
<?xml version="1.0"?>
<!DOCTYPE ead PUBLIC "-//Society of American Archivists/DTD
ead.dtd (Encoded Archival Description (EAD) Version 1.0)/EN" "ead.dtd">
<ead>
  <eadheader>
    <eadid>GBR/EAD/0016/TURNER</eadid>
    <filedesc>
      <titlemt>
        <titleproper>
          Papers of Dawson Turner (1775-1858), botanist, banker and antiquary
        </titleproper>
      </titlemt>
      <editionstmt>
        <edition>1.1</edition>
      </editionstmt>
      <publicationstmt>
        <publisher>Trinity College Library, Cambridge</publisher>
      </publicationstmt>
    </filedesc>
    <profiledesc>
      <creation>
        Ruth Bridgen and Jonathan Smith
        <date>17 April 2003</date>
      </creation>
      <language>
        <language>english</language>
        <language></language>
        <language></language>
      </language>
    </profiledesc>
  </eadheader>
```

Webový formulář, který je k dispozici na stránkách projektu Janus, provází katalogizátora jednotlivými částmi popisu. V první části se vytváří hlavička EAD se základními údaji o záznamu (název zdroje, katalogizátor, instituce, jazyk atd.). Po vyplnění údajů přechází vždy katalogizátor tlačítkem *nex*“ na další část formuláře. Druhá část obsahuje základní údaje o popisovaném archivu či sbírce. Katalogizátor vybírá především úroveň popisu a následně vyplňuje informace o názvu sbírky, časovém rozmezí, žánru, původci sbírky, případně odkaz na digitální obraz sbírky,... Třetí část formuláře tvoří prvky <bioghist>, <scopecontent> a <arrangement>, které mohou obsahovat podrobnější popis obsahu sbírky či archivu. Následuje popis, který shrnuje informace o provenienci sbírky, historii jejího uložení, způsobech rozšiřování a jejího ocenění. Pátá část se věnuje možnostem a omezením přístupu ke sbírce a jejím využití. Další část obsahuje poznámky a informace o souvisejících materiálech. Formulář zakončuje obrazovka s přístupovými prvky, které by měly být

vytvářeny podle řízených slovníků, tezaurů a obdobných pomůcek. Jde o jména osob, rodin, korporací, geografická jména a předmětová hesla.

Po potvrzení tlačítka *submi*“ se do okna Internetového prohlížeče vygeneruje metadatový záznam ve formátu EAD, který je možné zkopírovat a případně dále upravit.

První část webového formuláře pro vytváření záznamů ve formátu EAD, který je na svých stránkách Janus zpřístupňuje, je zobrazena na obrázku 11 (Janus, 2006c).

Obr. 11: Část webového formuláře pro vyplnění údajů pro hlavičku EAD

About the EAD file

eadid:

Title of EAD file:

edition:

Author of EAD file:

Date:

Published by:

Languages of EAD file:

Revision notes:

Pro katalogizátory se ale nabízí komfortnější možnost vytváření EAD záznamů v souladu s ISAD(G) pomocí databáze MS Access nazvané Cantab, kterou vyvinuli archiváři z Cambridge (Janus, 2006b).

6.3.3 Vyhledávání

Pro vyhledávání v bázi Janus se nabízí jednoduché vyhledávání, pokročilé vyhledávání a listování ve vybraném katalogu nebo v rejstřících.

Formulář pokročilého vyhledávání dovoluje použít pro spojení vyhledávaných výrazů nebo frází operátory AND, OR, AND NOT. Vyhledávání může probíhat s omezením na určitý prvek i s filtrem na vybraný archiv. Pod formulářem se zobrazují stručné výsledky hledání. Přes link je možné vstoupit na detailní náhled vybraného záznamu.

Při listování v katalozích nejdříve uživatel musí vybrat instituci, v jejímž archivu či sbírce chce vyhledávat. Následně se mu zobrazí základní informace o instituci, kontakty, případně otvírací doba. Pod záznamem instituce je uveden seznam sbírek a archivů, které daná instituce zpřístupňuje. Seznam může být v řadě případů velmi obsáhlý.

V případě, že instituce uplatňuje popis na více úrovních, je možné přes vybrané záznamy vstupovat na nižší a nižší úroveň popisu. Počet hierarchických úrovní záleží na každé instituci a také na každé sbírce.

Při sestupování na nižší a nižší úroveň se v levé části okna prohlížeče zobrazuje seznam sbírek vždy o úroveň vyšší. Je tedy možné snadno přecházet mezi sbírkami.

Při listování v rejstřících si nejdříve uživatel volí typ rejstříku. K dispozici má rejstřík předmětových hesel, osobních jmen, jmen rodin a rodů, jmen korporací a geografických názvů. Pak si vybírá konkrétní heslo. Spolu s heslem se uživateli zobrazuje seznam záznamů, které dané heslo obsahují. Přes odkaz před záznamem je možné zobrazit detail záznamu.

V levé horní části okna se zobrazuje pomůcka pro rychlejší navigaci. Tentokrát jde o hesla z rejstříku, která předchází vybranému výrazu a která za ním následují. Opět představují aktivní odkazy, které uživatele rychle přenesou na nový výraz a s ním související záznamy.

6.3.4 Shrnutí informací k projektu Janus

Janus je úspěšně se rozvíjející projekt. Nezaměřuje se pouze na historické sbírky o výchově a vzdělávání, které vlastní Univerzita v Cambridge, ale díky aktivní účasti institucí v celé oblasti kolem Cambridge obsahuje dokumenty z řady oborů. Sbírký jsou velmi různorodé i díky jednotlivcům, kteří sbírky původně shromažďovali. Je možné na nich sledovat vývoj celé britské kultury, vědy a umění. Lze v nich najít dokumenty k politickým otázkám či k tématům z lékařství a medicíny (Janus, 2006a).

Jde bezpochyby o zajímavý projekt, který je otevřený i ke spolupráci s jinými, již existujícími, projekty. Metadatové záznamy ve formátu EAD výrazně nemění a neupravuje oproti původnímu formátu, takže je pro něj velmi snadné přebírat záznamy jiných institucí v tomto formátu.

Díky nástrojům pro tvorbu, které poskytuje spolupracujícím institucím, a následným kontrolním a ladícím mechanismům a validacím, jsou záznamy do velké míry zpracované jednotně.

Zajímavá je na nich především možnost uspořádat záznamy do řady hierarchických úrovní. Software, se kterým pracuje koncový uživatel, je pro zobrazování těchto hierarchických úrovní dobře přizpůsoben.

7 ZÁVĚR

Jak z uvedených příkladů vyplývá, metadata se v současnosti používají v nejrůznějších oblastech lidské činnosti a k popisu různých objektů. Metadata dokáží popsat tradiční dokumenty stejně jako elektronické zdroje a objekty reálného světa vůbec. Pro každý účel jsou metadata poněkud odlišná.

Pohledy na metadata se mohou různit. Jinak vnímá metadata knihovník, jinak programátor a jinak např. obchodník. Přesto mají všechny pohledy na metadata něco společného. Jejich cílem je vždy napomoci vyhledávání, čitelnosti a opakovanému využití zdrojů. Obvykle je také důležité, aby bylo možné metadatové záznamy vyměňovat a sdílet.

V současnosti hrají metadata významnou roli v oblasti informačních a knihovnických institucí v popisu elektronických zdrojů a objektů a při budování digitálních archivů a knihoven. Využívána jsou také v oblastech vědy a výzkumu, vzdělávání a obchodování.

Metadata by měla pomoci vytvořit sémantický web. Spolu se strukturou RDF by měla uspořádat existující zdroje na webu a zkvalitnit vyhledávání. Také mohou zajistit kontrolu přístupu ke zdrojům, které jsou chráněny autorskými právy a určitými omezeními. Problémem, který možná v budoucnosti vyřeší používání vícejazyčných slovníků, je množství jazyků, ve kterých jsou metadatové záznamy vytvářeny.

Běžný člověk se s metadaty setkává velmi často, ale neuvědomuje si, že o metadata jde. V řadě systémů, kde jsou metadata použita, jsou záznamy skryté a pomocí transformací a uživatelského softwaru se zobrazují ve srozumitelné a uživatelsky přívětivé formě.

7.1 Popisné jednotky

Analyzovala jsem pouze vybrané projekty a systémy, které mě něčím zaujaly. Přesto mě překvapilo, jak různé jsou objekty, které tyto projekty a systémy považují za popisné jednotky. Očekávala jsem, že většinou budou popisnou jednotkou elektronické texty, které jsou obvyklé na Internetu. Předpokládala jsem, že hlavní problémy budou spočívat v rozhodnutí, zda popisovat komplex webových stránek

nebo jednotlivé stránky a to pouze v případech, kdy budou celé komplexy odkazovat na texty se stejnou tematikou.

Volba popisné jednotky se však ukázala jako mnohem složitější. Často projekty a systémy popisují objekty na více úrovních a snaží se tyto úrovně vzájemně provázat. Nejde jen o úroveň sbírka – jednotka, ale často je hierarchie mnohem propracovanější. Je pak možné popisovat a provazovat objekty na úrovni sbírky - části sbírky vymezené určitým způsobem - jednoho zdroje – části zdroje (jeden soubor, jedna kapitola, jeden list, obrázek v textu apod.). Jako příklad systému s řadou navazujících hierarchických úrovní může sloužit právě posledně popisovaný projekt Janus.

Velmi zajímavou se ukázala problematika popisu korpusů, které jsou často složeny z několika textů. Zejména v těchto případech se uplatňuje popisování celého korpusu i jednotlivých textů, které jsou v korpusu obsažené. V případě standardu CES, který využívá k bibliografickému popisu hlavičky TEI, má záznam několik úrovní.

V pojetí popisných jednotek nepanuje a ani nemůže panovat jednotnost a shoda. Z analyzovaných projektů vyplývá, že stanovení popisné jednotky je vždy závislé na cílech a očekávaných službách systémů. Není možné např. stanovit jako popisnou jednotku pro všechny systémy jednotlivý dokument, když jsou služby jednoho systému založeny na zpřístupňování záznamů o celých sbírkách dokumentů.

Nejednotnost popisných jednotek znemožňuje spolupráci některých projektů a systémů. Vylučuje možnost sdílení jejich záznamů a značně ztěžuje také konverzi záznamů. Myslím však, že nemá příliš velký význam sdílet a vyměňovat záznamy některých zcela odlišných systémů, které jsou zaměřeny nejen na různé obory, ale také na zcela odlišné uživatelské skupiny.

Význam by mělo propojení všech systémů a projektů na zcela jiné úrovni. Jistě by se i uživateli líbilo rozhraní, kde by mohl vyhledávat na nejvyšší úrovni nejdříve v databázi či archivu všech archivů, sbírek apod. Po volbě příslušného archivu by se zaměřil na jeden vybraný archiv. V rámci něho by mohl vyhledávat na nižší úrovni, např. na úrovni jednotlivých tematických fondů a sbírek daného archivu. Pak by se ponořil na úroveň ještě o něco nižší, na úroveň jednotlivých dokumentů, archiválií, textů apod. Po zvolení jednoho díla by mohl prohledávat dílo samotné, např. by se zaměřil na určitou kapitolu, obrázek, tabulku apod.

Vyhledávání by muselo být na každé úrovni jiné. Na nejvyšší úrovni by bylo zcela obecné, uživatel by nebyly k dispozici užitečné nástroje typu boolovské a proximitní operátory, řízené slovníky ani tezaury. Nebylo by to možné vzhledem k různorodosti systémů. Bylo by naopak nutné vyřešit dotazovací a vyhodnocovací systém, který by musel obsahovat konverzní tabulky a mechanismy pro vyhledávání. Na nižších úrovních by mohlo být vyhledávání sofistikovanější a nabízelo by se více možností, jak dotaz položit.

Spolupráce systémů, které nepoužívají stejné standardy, je složitá. Nezbytností jsou konverzní tabulky, které řeší především problematiku nestejných metadatových formátů a nejednotných používaných řízených slovníků, předmětových třídění a souborů autorit. Nikdy se asi nepodaří všechny tyto nástroje sjednotit, ale pravděpodobně bude stále sílit snaha o sjednocování a spolupráci a tvorbu nástrojů pro konverzi a převody nejen samotných formátů ale i těchto nástrojů.

7.2 Metadatové formáty

Analýza jasně ukázala, že není možné vytvořit jediný metadatový formát, který by vyhovoval všem projektům a systémům. Nejen každá komunita odborníků, ale dokonce i každá skupina v rámci jedné komunity, má často jiné požadavky na popis. Může to být i kvůli odlišným účelům, pro které chtějí vytvořené záznamy používat.

Snahy o vytvoření jednotného systému, který bude sloužit alespoň pro jednu oblast lidské činnosti se ukazují jako užitečné, ale ne zcela realizovatelné. Některé systémy sice přejímají původní metadatové formáty bez úprav, ale malou i když významnou změnou může být např. použití vlastního tezauru nebo předmětového třídění. I použití podobných nástrojů lze považovat za odchylku od původního formátu, která může při sdílení nebo výměně záznamů působit neshody.

Výjimkou nejsou projekty a systémy, které sice staví na vybraném formátu, ale, protože původní prvky nepostačují jejich potřebám, přidávají k formátu nové vlastní prvky, nebo alespoň upřesňují jejich význam pomocí kvalifikátorů. Často tak vznikají nové formáty, které opět mohou sloužit jako výchozí formáty pro jiné projekty a systémy. Jako příklad lze uvést australský formát AGLS, který přejala řada dalších australských projektů a systémů.

Některé projekty a systémy vznikají jako národní. Obvykle je pak snaha vycházet ze stávajících katalogizačních pravidel, standardů a postupů. Některé projekty a systémy se snaží přispět k bibliografické kontrole a zachování kulturního dědictví, přispívají také ke tvorbě národních bibliografií. Mezi takové projekty patří i české projekty WebArchiv a ANL FULL. Zatímco dnes pozastavený ANL FULL se zaměřuje pouze na články z českých novin a časopisů, WebArchiv sbírá různé zdroje obsažené na Internetu v doméně .cz. WebArchiv sice zpřístupňuje formulář pro tvorbu metadatových záznamů a pro vybrané zdroje vytváří ručně metadatové záznamy a záznamy ve formátu MARC21, ale jeho hlavním cílem je automatický sběr metadat. Modifikace formátu Dublin Core v případě projektu WebArchiv spočívá pouze v použití některých českých tezaurů, řízených slovníků a předmětových třídění. Pro projekt ANL FULL metadatové prvky Dublin Core nedostačovaly. Bylo nutné přidat další prvky, které by odpovídaly popisu článků. Zdánlivě podobné projekty, které se vyvíjejí v jedné zemi, a přesto je možné mezi nimi najít tolik odlišností. Už ani pro ně není možné použít jednotný formát. O kolik těžší by bylo tvořit jednotný metadatový formát, který by vyhovoval všem projektům a systémům!

Využití konverzních tabulek pro výměnu a sdílení záznamů, které nepoužívají stejné formáty, je pouze částečné řešení. Vytvořit kvalitní konverzní tabulky je velmi náročné. Formáty nemusí mít odpovídající prvky, některé prvky mohou chybět, jiné přebývat. Dalším problémem je obsah prvků. V jednom formátu může být obsah prvků vybírán z řízených slovníků, tezaurů či předmětových třídění, další formát může mít v prvcích obsažen volný text. Na tohle vše musí konverzní tabulka najít řešení.

V oblasti popisných metadatových formátů se situace relativně stabilizuje. Nejvýznamnější popisné metadatové formáty pro nejbližší roky již s největší pravděpodobností vznikly a v současnosti lze spíše očekávat systémy a projekty na ně navazující. Lze usuzovat, že pro každý z těchto systémů bude vybraný formát určitým způsobem modifikován, případně přinejmenším přijata určitá doporučení pro uvádění povinných prvků a pro využívání definovaných řízených rejstříků, autoritníchází, tezaurů apod. Vzhledem k velké rozmanitosti systémů není možné stanovit pro všechny jednotná pravidla. Každý popis bude asi vždy do jisté míry jedinečný, přestože budou stavět na stejném metadatovém formátu.

Většina projektů se snaží zajistit spolupráci maximálního množství kulturních institucí. Kromě jednotného vyhledávacího portálu se tak uživatelům nabízí informace o řadě zdrojů na jednom místě, ve kterých je možné vyhledávat jednotným způsobem. Záleží na způsobu spolupráce, který řídící instituce volí. Není vždy nutné přijmout striktní pravidla popisu stanovená centrálně. Obvykle je možné, aby spolupracující instituce využívaly vlastní řízené rejstříky, tezaury a podobné pomůcky. V některých případech dokonce není nutné ani využívat pro popis stejný metadatový formát, ale data mohou být transformována. Je-li při spolupráci uplatněn distribuovaný model, má každá instituce větší kontrolu nad vlastními daty.

Je také vysoce pravděpodobné, že v souvislosti se sdílením a výměnou záznamů v různých popisných metadatových formátech bude růst význam strukturálních metadatových formátů. Jejich výhodou je sdružení všech typů informací, které se váží k jednomu objektu. Databáze či repozitář záznamů METS je tedy výborný zdroj informací o objektech bez ohledu na popisný metadatový formát. Následně jde „pouze“ o software, který musí poskytnout dostatečně sofistikované nástroje pro vyhledávání, zobrazování a řízený přístup k vyhledaným zdrojům. Pro popis elektronických zdrojů ve světě knihoven, archivů a dalších paměťových institucích se nejvíce prosazuje strukturální formát METS.

POUŽITÉ ZKRATKY

AACR - Anglo-American Cataloging Rules
AAM - American Association of Museums
AASLH - American Association for State and Local History
AAT - Art & Architecture Thesaurus
ACL - Association for Computational Linguistics
AGIFT - Australian Government Interactive Functions Thesaurus
AGLS - Australian Government Locator Service
AHDS - Arts and Humanities Data Service
ACH - Association for Computers in Humanities
ALLC - Association for Literary and Linguistic Computing
ANSI - American National Standards Institute
ANZLIC - Australia and New Zealand Land Information Council
ASC - Association of Systematics Collections
ATED - Australian Thesaurus of Education Descriptors
BEP - Business Entry Point
BFAP - Berkeley Finding Aid Project
BIC - Book Industry Communication
BNC - British National Corpus
CAIE - Committee on Archival Information Exchange
CCS - Cascading Style Sheets
CEN – European Committee for Standardization
CES - Corpus Encoding Standard
CIHM - Canadian Institute for Historical Microreproductions
CIMI - Computer Interchange of Museum Information
CIS - Common Information System
CISAC – International Confederation of Societies of Authors and Composers
CNRS - Centre National de la Recherche Scientifique
CoBRA - Computerised Bibliographic Actions
CORC - Cooperative Online Resource Catalog
CURL - Consortium of University Research Libraries
DC – Dublin Core
DCMI – Dublin Core Metadata Initiative
DESIRE - Development of a European Service for Information on Research and Education
DNER - Distributed National Electronic Resource
DOI - Digital Object Identifier
DSSSL - Document Style Semantics and Specification Language
DSTC - Distributed Systems Technology Centre
DTD - Document Type Definition
EAD - Encoding Archival Description
EAGLES - Expert Advisory Group on Language Engineering Standards
ECI - European Corpus Initiative
ECO - Early Canada Online
EDT-MS - Electronic Theses and Dissertations
eLib - Electronic Libraries
ELRA - European Language Resources Association
ERIC - Educational Resources Information Center
ESRC - Economic and Research Council

FRANAR -Functional Requirements and Numbering for Authority Records
 FRBR - Functional Requirements for Bibliographic Record
 FRSAR - Functional Requirements of Subject Authority Record
 GILS - Government Information Locator Service
 HASSET - Humanities And Social Sciences Electronic Thesaurus
 HTML – Hyper Text Markup Language
 CHIO - Cultural Heritage Information Online
 IAFA - Internet Anonymous Ftp Archives
 IETF - Internet Engineering Task Force
 IFLA - International Federation of Library Associations and Institutions
 IFPI – International Federation of Phonographical Industry
 ILRT - Institute of Learning and Research Technology, University of Bristol
 ILS - Integrated Library System
 INDECS - Interoperability of Data in E-Commerce Systems
 ISAAR - International Standard Archival Authority Record for Corporate Bodies, Persons, and Families
 ISAD - International Standard of Archival Description
 ISADN - International Standard Authority Data Number
 ISAN - International Standard Audiovisual Number
 ISBD (ER) – International Standard Bibliographic Description (Electronic Resources)
 ISBN – International Standard Book Number
 ISMN - International Standard Music Number
 ISO – International Organization for Standardization
 ISRC - International Standard Record Code
 ISRN - international Standard Record Number
 ISSN – International Standard Serial Number
 ISWC - International Standard Musical Work Code
 JISC - Joint Information Systems Committee
 JST - Japan Science and Technology Corporation
 LCC – Library of Congress Classification
 LCSH - Library of Congress Subject Heading
 MAB - Maschinellen Austauschformat für Bibliotheken
 MARC – Machine Readable Cataloging Record
 MASTER - Manuscript Access through Standards for Electronic Records
 MATE - Multilevel Annotation, Tools Engineering
 MCN - Museum Consortium Network
 METS - Metadata Encoding and Transmission Standard
 MIMAS - Manchester Information & Associated Services
 MMSB - Memoriae Mundi Series Bohemica
 Multext - Multilingual Text Tools and Corpora
 MUSE - Digital Media Communication System
 NAA - National Archives of Australia
 NCA - National Council on Archives
 NCSA - National Centre for Supercomputing Applications
 NDLTD - Networked Digital Library of Theses and Dissertations
 NEDLIB - Networked European Depository Library
 NII - National Information Infrastructure
 NISO - National Information Standard Organization
 NOIE - National Office for the Information Economy
 NSF - National Science Foundation
 NWA - Nordic Web Archive,
 OAI - Open Archive Initiative

OAI-PMH - Open Archives Initiative – Protocol for Metadata Harvesting
OCLC - Online Computer Library Center
OCR - Optical Character Recognition
ODL - Oxford Digital Library
OPAC – Online Public Access Catalog
OTA - Oxford Text Archive
OTASH - Oxford Text Archive Subject Headings
PAROLE - Preparatory Action for Linguistic Resources Organisation for Language Engineering
PICS - Platform for Internet Content Selection
RDA - Resource Description and Access
RDF - Resource Description Framework
RFC – Request for Comments
RLG – Research Library Group
ROADS - Resource Organisation and Discovery in Subject-based Services
SAA - Society of American Archivists
SDLIP - Simple Digital Library Interoperability Protocol
SGML – Standard Generalized Markup Language
SICI - Serial Item and Contribution Identifier Standard
SOSIG - Social Science Information Gateway
SPAG - Standards and Practices Action Group
SURA - Southeastern Universities Research Association
TEI - Text Encoding Initiative
TGN - Thesaurus of Geographic names
UKAT – United Kingdom Archival Thesaurus
UKOLN – UK Office for Library and Information Networking
UMI - University Microfilms
UNESCO - United Nations Educational, Scientific and Cultural Organisation
UNIMARC – Universal Machine Readable Cataloging Record
URI - Uniform Resource Identifier
URL – Uniform Resource Locator
URN - Uniform Resource Names
USMARC – United States Machine Readable Cataloging Record
VOCED – Vocational Education and Training
VRA – Visual Resources Association
W3C - World Wide Web Consortium
WERA – Web Archive Access
WWW – World Wide Web
XHTML – Extensible HyperText Markup Language
XML - eXtensible Markup Language
XSL - eXtensible Stylesheet Language
XSLT – XSL Transformations

SEZNAM POUŽITÉ LITERATURY

- *About the TGN* [online]. J. Paul Getty Trust, revised 6 December 2006 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.getty.edu/research/tools/vocabulary/tgn/about.html>>.
- ANDĚROVÁ, Ivana. 2000. Programový projekt MK ČR "Souborná databáze Kooperačního systému článkové bibliografie : optimalizace integrace a správy heterogenních dat". *Ikaros* [online]. 2000, roč. 4, č. 10 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.ikaros.cz/node/3976>>. ISSN 1212-5075.
- ANDĚROVÁ, Ivana. 2002a. Báze ANL FULL v systému TOPIC. In *Inforum*, 22. května 2002, Praha. Praha, 2002 [cit. 2007-08-31]. Dostupný také z WWW: <<http://www.inforum.cz/inforum2002/prednaska26.htm>>.
- ANDĚROVÁ, Ivana. 2002b. Kooperační systém článkové bibliografie - KOSABI (Vývoj a současný stav metodiky zpracování, zpřístupnění, organizace kooperace, perspektivy). In *Knihovny současnosti 2002. Sborník z 10. konference, konané ve dnech 24.–26. září 2002 v Seči u Chrudimi*. Brno : Sdružení knihoven ČR, 2002. s. 223-255. Dostupný také z WWW: <<http://full.nkp.cz/nkdb/docs/KOSABId.ppt>>. ISBN 80-86249-18-2.
- ANDĚROVÁ, Ivana. 2002c. *RE: ANL FULL* [elektronická pošta]. Message to: Veronika Vozková. 2002-11-11 12:33 [cit. 2002-11-11]. Osobní komunikace.
- ANDĚROVÁ, Ivana. 2004. *Kooperační systém článkové bibliografie - KOSABI , management KOSABI, Článkové informace v ČR* [online]. Praha : Národní knihovna ČR, 30.12.2004, aktualizace jaro 2007 [cit. 2007-08-31]. Dostupný z WWW: <<http://full.nkp.cz/nkdb/docs/mngkkochar.htm>>.
- ANDĚROVÁ, Ivana. 2005a. *Server full.nkp.cz, báze ANL FULL – stav v r. 2004 a 2005. Perspektivy. Návrh vývoje* [online]. Praha : Národní knihovna ČR, říjen 2005, aktualizace září 2006 [cit. 2007-08-31]. Dostupný z WWW: <http://full.nkp.cz/nkdb/docs/full_persp.htm>.
- ANDĚROVÁ, Ivana. 2005b. *Souborná databáze Kooperačního systému článkové bibliografie - optimalizace integrace a správy heterogenních dat : Souhrnná zpráva za rok 2001* [online]. Praha : Národní knihovna ČR, leden 2005 [cit. 2007-08-31]. Dostupný z WWW: <http://full.nkp.cz/nkdb/docs/soubor_final.htm>.
- ANDĚROVÁ, Ivana. 2005c. *Souborná databáze Kooperačního systému článkové bibliografie - optimalizace integrace a správy heterogenních dat : Závěrečná zpráva za rok 2000-2004* [online]. Praha : Národní knihovna ČR, leden 2005 [cit. 2007-08-31]. Dostupný z WWW: <http://full.nkp.cz/nkdb/docs/soubor_final.htm>. ISSN 1212-5075.
- Archives Hub. 2000a. *Contributors* [online]. Manchester : University of Manchester, c2000-2007, last updated May 11, 2007 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.archiveshub.ac.uk/repositories.shtml>>.
- Archives Hub. 2000b. *Creating access points* [online]. Manchester : University of Manchester, c2000-2007, last updated November 21, 2006 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.archiveshub.ac.uk/arch/access.shtml>>.
- Archives Hub. 2000c. *EAD 2002 Online Template* [online]. Manchester : University of Manchester, c2000-2006, last updated September 25, 2006 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.archiveshub.ac.uk/eadform2002.html>>.
- Archives Hub. 2000d. *Introduction to the Archives Hub* [online]. Manchester : University of Manchester, c2000-2007, last updated November 14, 2006 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.archiveshub.ac.uk/introduction.shtml>>.
- Archives Hub. 2000e. *ISAD(G) to EAD example* [online]. Manchester : University of Manchester, c2000-2007, last updated May 8, 2007 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.archiveshub.ac.uk/sample.shtml>>.

- Archives Hub. 2000e. *Sample EAD record* [online]. Manchester : University of Manchester, c2000-2007, last updated November 20, 2006 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.archiveshub.ac.uk/arch/sample.shtml>>.
- Archives Hub. 2000g. *Technology* [online]. Manchester : University of Manchester, c2000-2007, last updated November 21, 2006 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.archiveshub.ac.uk/technology.shtml>>.
- Archives Hub. 2002a. *Archive Hub workbook* [online]. Manchester : University of Manchester, January 2002, last updated February 2004 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.archiveshub.ac.uk/ahwkbk.html>>.
- Archives Hub. 2002b. *Introducing the Archives Hub* [online]. Manchester : University of Manchester, June 2002 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.archiveshub.ac.uk/ah101.html>>.
- Asociace knihoven vysokých škol České republiky. 2004. *Pracovní komise pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací : základní informace* [online]. 2004-06-11, aktualizace 13.6.2007 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.evskp.cz/>>.
- BARTOŠEK, Miroslav. 1999. Vyhledávání v Internetu a Dublin Core. *Zpravodaj ÚVT MU*. 1999, roč. 9, č. 4, s.1-4. ISSN 1212-0901.
- BEARMAN, David, et al. 1999. a Common Model to Support Interoperable Metadata : Progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities. *D-Lib Magazine* [online]. January 1999, vol. 5, no. 1 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.dlib.org/dlib/january99/bearman/01bearman.html>>. ISSN 1082-9873.
- BRATKOVÁ, Eva. 1999. Metadata jako nový nástroj pro komunikaci webovských informačních zdrojů. *Národní knihovna*. 1999, roč. 10, č. 4, s. 178-194. ISSN 0862-7487.
- BRATKOVÁ, Eva. 2001. *Informační entity, jejich znaky a vztahy : podle modelu FRBR : učební text pro posluchače ÚISK*. Praha : Ústav informačních studií a knihovnictví FF UK, 2001.
- BRATKOVÁ, Eva. 2002. Elektronické disertace a projekty a systémy jejich zpřístupňování. *Národní knihovna*. 2002, roč. 13, č. 4, s. 268-287. Dostupný také z WWW: <<http://full.nkp.cz/nkkp/Nkkp0204/0204268.html>>. ISSN 1214-0678.
- BRATKOVÁ, Eva. 2003. Principy, aktéři, modely, postupy a technologie v systémech elektronických disertací. In *Inforum 2003 : 9. konference o profesionálních informačních zdrojích, Praha, Vysoká škola ekonomická, 27.-29.5.2003* [online]. Praha : Vysoká škola ekonomická, 2003 [cit. 2007-08-31]. Dostupný z WWW: <http://www.inforum.cz/inforum2003/prispevky/Bratkova_Eva.pdf>.
- BURNARD, Lou et al. 1996. *a syntax for Dublin Core metadata : recommendations from the second meta workshop* [online]. Last modified 19-Apr-1996 [cit. 2007-08-31]. Dostupný z WWW: <<http://tiger.uic.edu/~cmsmcq/tech/metadata.syntax.html>>.
- BURNARD, Lou; DRISCOLL, Matthew; RAHTZ, Sebastian. 2005. Digital Texts, XML, and TEI. In *TEI training, Sofia, October 2005* [online]. Text Encoding Initiative, [2005] [cit. 2007-08-31]. Dostupný z WWW: <<http://www.tei-c.org/Talks/2005/Sofia/talk-intro.pdf>>.
- BURNARD, Lou; LIGHT, Richard. 1998. Three SGML metadata formats : TEI, EAD, and CIML : a study for BIBLINK. *Work Package 1.1* [online]. December 1996, last update 14 May 1998 [cit. 2007-08-31]. Dostupný z WWW: <<http://hosted.ukoln.ac.uk/biblink/wp1/sgml/>>.
- BURNARD, Lou; ROBINSON, Peter. 1999. *Towards a European Standard for Manuscript Description : the MASTER project* [online]. 12 Jul 99 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.tei-c.org.uk/Master/Hermes/front.htm>>.
- BURNARD, Lou (ed.). 2001. Reference Manual for the MASTER Document Type Definition : discussion draft [online]. Oxford University Computing Service, revised 6 Jan 2001 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.tei-c.org.uk/Master/Reference/oldindex.html>>.

- BYRUM, John D.; MADISON, Olivia M. A. 2000. Reflections on the goals, concepts and recommendations of the IFLA Study on functional requirements of bibliographic records. In *FRBR (Functional requirements for bibliographic records) Seminar - Florence, 27-28 January 2000* [online]. 2000, last revision 2000-02-06 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.aib.it/aib/sezioni/toscana/conf/frbr/byrmadis.htm>>.
- CARPENTER, Leona. 2003. *OAI for beginners – the Open Archives Forum online tutorial* [online]. Last modified 14 Oct 2003 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.oaforum.org/tutorial/index.php>>.
- CELBOVÁ, Ludmila. 2001. Informace o projektu registrace domácích internetových zdrojů nově na serveru WebArchiv. *Ikaros* [online]. 2001, č. 5 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.ikaros.cz/node/754>>. ISSN 1212-5075.
- CELBOVÁ, Ludmila. 2002. Balance pilotního projektu WebArchiv anebo Co bude dál? *Ikaros* [online]. 2002, č. 3 [cit. 2007-08-31]. Dostupný na WWW: <<http://www.ikaros.cz/node/918>>. ISSN 1212-5075.
- CELBOVÁ, Ludmila; Žabička, Petr. 2004. Archivace internetových zdrojů v prostředí knihoven. In *Sbírký v počítači. Muzea, archivy a knihovny v éře digitálních informací*. Praha: NTM, 2004. Dostupný také z WWW: <<http://www.webarchiv.cz/files/dokumenty/konference/ntm2004.pdf>>. ISBN 80-7037-131-5.
- CELBOVÁ, Ludmila. 2006. Ochrana a trvalé zpřístupnění webových zdrojů jako součásti národního kulturního dědictví : průběžná zpráva realizace projektu v roce 2006 [online]. Praha : Národní knihovna, 2006 [cit. 2007-08-31]. Dostupný z WWW: <http://www.webarchiv.cz/files/dokumenty/zpravy/zprava-VaV_2006-final.rtf>.
- Centre National de la Recherche Scientifique. 1996. *Multext : Multilingual Text Tools and Corpora* [online]. Verze 0.1. 22 April 1996 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.lpl.univ-aix.fr/projects/multext/>>.
- Centre National de la Recherche Scientifique. 2006. *Overview* [online]. [2006] [cit. 2007-08-31]. Dostupný z WWW: <<http://www2.cnrs.fr/en/8.htm>>.
- COLLIS, Paddy. 2001. *Archives Hub online template : EAD version 2002 for XML: guidelines, original version May 2001, last revised March 2006* [online]. Manchester : University of Manchester, revised March 2006 [cit. 2007-05-15]. Dostupný z WWW: <<http://www.archiveshub.ac.uk/arch/dcg0603.pdf>>.
- COVER, Robin. 2001. *Manuscript Access through Standards for Electronic Records (MASTER)* [online]. Organization for the Advancement of Structured Information Standard, February 10, 2001 [cit. 2007-08-27]. Dostupný z WWW: <<http://xml.coverpages.org/master.html>>.
- COYLE, Karen. 2005. Understanding Metadata and its Purpose. *Journal of Academic Librarianship*. 2005, vol. 31, n. 2, s. 160-163. ISSN 0099-1333.
- CULBERTSON, Becky. 2001. *Comparison of subject treatment of several metadata standards* [online]. La Jolla (CA) : University of California, September 25, 2001 [cit. 2007-08-31]. Dostupný z WWW: <<http://library.ucsd.edu/~becky/metadata.htm>>.
- CUNNINGHAM, Adrian. 1998. *Australian Government Locator : enabling seamless online access to government* [online]. Canberra : National Archives of Australia, date issued 1998-08-26 [cit. 2007-05-15]. Dostupný z WWW: <http://www.naa.gov.au/recordkeeping/gov_online/agls/Metadata_paper22sept98.html>.
- CUNNINGHAM, Adrian. 2003. *The AGLS Metadata Element Set goes national* [online]. Canberra : National Archives of Australia, created 2003, modified 2005-06-27 [cit. 2007-05-15]. Dostupný z WWW: <http://www.naa.gov.au/recordkeeping/gov_online/agls/enabling_online_access.html>.
- CVRČKOVÁ, Radomíra. 2000. Služba GILS jako nástroj pro řízení informačních zdrojů z oblasti státní správy USA. Praha, 2000. 128 s. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí diplomové práce PhDr. Eva Bratková.

- Česko. 2006. Zákon 216/2006 Sb., kterým se mění zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů. In *Sbírka zákonů České republiky*. 2006, částka 72, s. 2707-2726. Dostupný také z WWW: <<http://www.mvcr.cz/sbirka/2006/sb072-06.pdf>>.
- DAY, Michael. 2004. Preservation metadata. *Metadata applications and management, International Yearbook of Library and Information Management, 2003-2004* / G. E. Gorman and Daniel G. Dorner (eds.). London: Facet Publishing, 2004, s. 253-273. Preprint kapitoly dostupný z WWW: <<http://www.ukoln.ac.uk/metadata/publications/iylim-2003/>>. ISBN 1-85604-474-2 (Facet Publishing); ISBN 0-8108-4980-1 (Scarecrow Press).
- DCMI. 1995a. *DC-2006. Metadata for knowledge management and learning : International conference on Dublin Core and metadata applications, 3-6 October 2006* [online]. DCMI, c1995-2006 [cit. 2007-08-31]. Dostupný z WWW: <<http://dc2006.ucol.mx/>>
- DCMI. 1995b. *DC8 – Agenda with links to presentation* [online]. DCMI, c1995-2007 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.dublincore.org/workshops/dc8/agenda-resources.shtml>>.
- DCMI. 1995c. *DCMI frequently asked questions (FAQ)* [online]. DCMI, c1995-2007 [cit. 2007-08-31] [cit. 2004-04-05]. Dostupný z WWW: <<http://dublincore.org/resources/faq/>>.
- DCMI. 1995d. *DCMI working groups* [online]. DCMI, c1995-2007 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.dublincore.org/groups/>>.
- DCMI. 1995e. *Dublin Core Metadata Initiative Document Translations* [online]. DCMI, c1995-2007 [cit. 2007-08-31]. Dostupný z WWW: <<http://dublincore.org/resources/translations/>>.
- DCMI. 1995f. *Dublin Core Metadata Initiative home page* [online]. DCMI, c1995-2007, last updated 7 May 2007 [2007-08-31]. Dostupný z WWW: <<http://dublincore.org/index.shtml>>.
- DCMI. 1995g. *Principles and Operation of the Dublin Core Metadata Initiative* [online]. DCMI, c1995-2007 [cit. 2007-08-31]. Dostupný z WWW: <<http://dublincore.org/about/organization/>>.
- DCMI. 2000. *Dublin Core qualifiers* [online]. Dublin (OH) : Dublin Core Metadata Initiative, 2000-07-11 [cit. 2007-08-31]. Dostupný z WWW: <<http://purl.org/dc/documents/rec/dcmes-qualifiers-20000711.htm>>.
- DCMI. 2005. *DC-2005. Vocabularies in practice : International conference on Dublin Core and metadata applications, 12-15 September 2005, Madrid* [online]. Madrid : University Carlos III of Madrid, 2005-02-07 [cit. 2007-08-31]. Dostupný z WWW: <<http://dc2005.uc3m.es/index.asp/>>.
- DCMI. 2006a. *DCMI metadata terms* [online]. DCMI, 2006-12-18 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.dublincore.org/documents/dcmi-terms/>>.
- DCMI. 2006b. *Dublin Core metadata element set. Version 1.1* [online]. DCMI, 2006-12-18 [cit. 2007-08-09]. Dostupný z WWW: <<http://dublincore.org/documents/dces/>>.
- DROBÍKOVÁ, Barbora. 2002. Vývoj, směřování a trendy katalogizace za poslední čtyři roky: od FRBR až po revizi AACR2R v roce 2002. *Národní knihovna*. 2002, roč. 13, č. 3, s. 153–167. Dostupný také z WWW: <<http://full.nkp.cz/nkkrr/Nkkrr0203/0203153.html>>. ISSN 1214-0678.
- EAD. 1998. *EAD tag library for version 1.0. tag library* [online]. Chicago : Society of American Archivists, June 1998 [cit. 2007-05-15]. Dostupný z WWW: <<http://www.loc.gov/ead/tglib1998/tlhome.html>>.
- EAD. 2002a. *Desing principles for enhancements to EAD : December 2002* [online]. Washington (DC) : Library of Congress, [2002] [cit. 2005-05-15]. Dostupný z WWW: <<http://www.loc.gov/ead/eaddesgn.html>>.
- EAD. 2002b. *Development of the Encoded Archival Description Document Type Definition* [online]. Washington (DC) : Library of Congress, [December 2002], last update June 5, 2006 [cit. 2007-05-15]. Dostupný z WWW: <<http://lcweb.loc.gov/ead/eadback.html>>.

- EAD. 2002c *Encoded Archival Description Tag Library, Version 2002 : Appendix C: Encoded Examples* [online]. Chicago : Society of American Archivists, May 26, 2006 [cit. 2007-08-30]. Dostupný z WWW: < http://www.loc.gov/ead/tglib/appendix_c.html>.
- EAD. 2002d *Encoded Archival Description Tag Library, Version 2002 : EAD elements by tag name* [online]. Chicago : Society of American Archivists, May 26, 2006 [cit. 2007-08-31]. Dostupný z WWW: <http://www.loc.gov/ead/tglib/element_index.html>.
- EAD. 2002e. *Encoding Archival Description Tag Library, Version 2002 : EAD Technical Document No. 2* [online]. Chicago : Society of American Archivists, c2002 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.loc.gov/ead/tglib/index.html>>. ISBN 1-931666-00-8.
- EAD. 2006. *Development of the Encoded Archival Description DTD* [online]. Washington (DC) : Library of Congress, June 5 2006 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.loc.gov/ead/eadddev.html>>.
- Expert Advisory Group on Language Engineering Standards. 2000. *The essentials of EAGLES* [online]. [2000] [cit. 2007-08-31]. Dostupný z WWW: <<http://www.ilc.cnr.it/EAGLES/intro.html>>.
- EDNA. 2002. *EdNA metadata standard v1.1* [online]. Commonwealth of Australia, September 2002 [cit. 2007-05-16]. Dostupný z WWW: <http://www.edna.edu.au/edna/webdav/site/myjahiasite/shared/edna_metadata.pdf>.
- EDNA. 2003a. *Metadata with EdNA Online* [online]. Version 2.2. Education.au, June 2003 [cit. 2003-08-15]. Dostupný z WWW: <<http://www.edna.edu.au/edna/go/pid/384>>. Uvedený odkaz byl funkční 15.8.2003, funkční odkaz se nepodařilo dohledat.
- EDNA. 2003b. *New generation EdNA Online goes live! : a collaborative enterprise* [online]. Commonwealth of Australia,, 2003 [cit. 2007-07-23]. Dostupný z WWW: <http://www.edna.edu.au/edna/webdav/site/myjahiasite/shared/site/new_gen_edna_online.pdf>.
- EDNA. 2006. *Harvesting overview* [online]. Version 1.0. Education.au, 2006 [cit. 2007-08-29]. Dostupný z WWW: <http://www.edna.edu.au/edna/webdav/site/myjahiasite/shared/edna2_harvesting.doc>.
- EDNA. 2007a. *About EdNA : Australia's free online network for educators* [online]. Commonwealth of Australia, c2007, last modified Wednesday, 27 June 2007 [cit. 2007-07-23]. Dostupný z WWW: <<http://www.edna.edu.au/edna/go/about>>.
- EDNA. 2007b. *About EdNA services* [online]. Commonwealth of Australia, c2007, last modified Wednesday, 27 June 2007 [cit. 2007-07-23]. Dostupný z WWW: <<http://www.edna.edu.au/edna/go/about/services>>.
- EDNA. 2007c. *Metadata harvesting* [online]. Commonwealth of Australia, c2007, last modified Monday, 30 April 2007 [cit. 2007-07-23]. Dostupný z WWW: <http://www.edna.edu.au/edna/go/resources/metadata/metadata_harvesting/>.
- EDNA. 2007d. *Strategic directions 2006-2008* [online]. Commonwealth of Australia, c2007, last modified Tuesday, 27 March 2007 [cit. 2007-07-23]. Dostupný z WWW: <http://www.edna.edu.au/edna/webdav/site/myjahiasite/shared/edna_plan_v1.pdf>.
- Elsevier. 2007. *About Scirus...* [online]. Elsevier, c2007 [cit. 2007-06-19]. Dostupný z WWW: <<http://www.scirus.com/srsapp/aboutus/>>.
- *Generátor metadat* [online]. 2001. Webarchiv, [2001] [cit. 2007-07-23]. Dostupný z WWW: <<http://webarchiv.cz/generator/dc.php>>.
- HAIGH, Susan. The Dublin Core Metadata Initiative. *Network notes* [online]. December 1999, no. 63 [cit. 2007-05-16]. Dostupný z WWW: <<http://www.collectionscanada.ca/9/1/p1-262-e.html>>. ISSN 1201-4338.
- HEERY, Rachel. 1996. Review of metadata formats. *Program*. October 1996, vol. 30, no. 4, s. 345-373. Dostupný také z WWW: <<http://www.ukoln.ac.uk/metadata/review.html>>. ISSN 0033-0337.

- HILL, Amanda. 2002. *RE: Archives Hub - Request for information* [elektronická pošta]. Message to: Veronika Vozkova. 2002-11-06 16:43 [cit. 2002-11-11]. Osobní komunikace.
- HILL, Amanda; HARRISON, John. 2005. Relinquishing Control: Developing the Distributed Archives Hub. *New Review of Information Networking*. 2005, vol. 11, no. 2, p. 159-170. Dostupný také z WWW: <http://www.archiveshub.ac.uk/relinquishing_control.pdf>. ISSN 1361-4576.
- HILL, Amanda. 2006. Opening Up the Archives: From Basement to Browser. *CILIP Rare Book and Special Collections newsletter*. [2006, no. 75.] Dostupný také z WWW: <http://www.archiveshub.ac.uk/CILIP_AHill.pdf>.
- HILLMANN, Dianne. 2004. An Introduction to Dublin Core. In *International Conference on Dublin Core and Metadata Application 2004, Shanghai, China, 11-14 October 2004* [online]. Shanghai, October 2004 [cit. 2007-05-16]. Dostupný z WWW: <http://www.dublincore.org/resources/training/dc-2004/english/DC-2004_Tutorial_1_en.pdf>.
- HORSÁK, Ondřej. 2005. *Struktura RDF pro metadata, její vývoj a perspektivy aplikace v digitálních systémech*. Praha, 2005. 114 s. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí diplomové práce PhDr. Eva Bratková.
- Humanities Text Initiative. 1993. *Guidelines for electronic text encoding and interchange (TEI P3)* [online]. Michigan : University of Michigan, [1993], modified 23.9.2000 [cit. 2007-05-16]. Dostupný z WWW: <<http://www.hti.umich.edu/t/tei/>>.
- HUTAŘ, Jan; MATĚJKA Lukáš; CELBOVÁ, Ludmila. 2006. Vyhledávání v archivu českých webových zdrojů. In *Inforum2006 : 12. ročník konference o profesionálních informačních zdrojích, Vysoká škola ekonomická, Praha, 23.-25. května 2006* [online]. Praha: Albertina icome Praha, 2006 [cit. 2007-05-16]. Dostupný z WWW: <http://www.inforum.cz/inforum2006/pdf/Hutar_Jan.pdf>. ISSN 1801-2213.
- CHACHRA, Vinod. 2006. a Union Catalog for Networked Digital Library for Theses and Dissertations – an update. *Proceedings of Fifth International Symposium on Electronic Theses and Dissertations, West Virginia University*. Provo (Utah) : NDLTD, 10 July 2006. Dostupný také z WWW: <<http://docs.ndltd.org:8080/dspace/handle/2340/196>>.
- CHAPESIUK, Ron. 1999. Organizing the Internet : the "core" of the challenge. *American Libraries*. Jan 1999, vol. 30, no. 1, s. 60-63. ISSN 0002-9769.
- IDE, Nancy. 1996. *Corpus Encoding Standard* [online]. Version 1.4. Last modified 14 October 1996 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.lpl.univ-aix.fr/projects/multext/CES/CES1.html>>.
- IDE, Nancy. 2002. *Re: CES - Request for information* [elektronická pošta]. Message to: Veronika Vozkova. 2002-11-04 18:03 [cit. 2002-11-11]. Osobní komunikace.
- *ISAD(G) : General International Standard Archival : Adopted by the Committee on Descriptive Standard, Stockholm, Sweden, 19-22 September 1999. Description*. 2nd ed. Ottawa : International Council on Archives, 2000. Dostupný také z WWW: <http://www.ica.org/biblio/cds/isad_g_2e.pdf>. ISBN 0-9696035-5-X.
- *ISO 15839:2003 : Information and documentation – The Dublin Core metadata element set* [online]. Geneva : ISO, [1993] [cit. 2007-08-31]. Dostupný z WWW: <<http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=37629&ICS1=35&ICS2=240&ICS3=30>>.
- Janus. 2006a. *About Janus* [online]. [2006] [cit. 2007-05-18]. Dostupný z WWW: <<http://janus.lib.cam.ac.uk/db/node.xsp?id=Webpages/Public/about>>.
- Janus. 2006b. *Cantab* [online]. [2006] [cit. 2007-05-18]. Dostupný z WWW: <<http://janus.lib.cam.ac.uk/db/node.xsp?id=Webpages%2FPrivate%2Fcontent%2Fcantab>>.
- Janus. 2006c. *Collection-level descriptions* [online]. [2006] [cit. 2007-05-18]. Dostupný z WWW: <<http://janus.lib.cam.ac.uk/db/node.xsp?id=Webpages%2FPrivate%2Fcontent%2Fcollection>>.

- Janus. 2006d. *EAD elements and the stylesheet* [online]. [2006] [cit. 2007-05-18]. Dostupný z WWW: <<http://janus.lib.cam.ac.uk/db/node.xsp?id=Webpages%2FPublic%2Flinks%2Fstylesheet>>.
- Janus. 2006e. *Welcome to first-time users* [online]. [2006] [cit. 2007-05-18]. Dostupný z WWW: <http://janus.lib.cam.ac.uk/db/node.xsp?id=Webpages%2FPublic%2Fnew_users>.
- Janus. 2006f. *What the "tidyer" does* [online]. [2006] [cit. 2007-05-18]. Dostupný z WWW: <<http://janus.lib.cam.ac.uk/db/node.xsp?id=Webpages%2FPublic%2Fabout%2Ftidyer>>.
- JELÍNKOVÁ, Lenka. 2005. Popis elektronických online zdrojů v rámci projektu WebArchiv. In *Informační setkání u příležitosti zpřístupnění digitálního archivu, Národní knihovna ČR, 6.12.2005* [online]. WebArchiv, 2005 [cit. 2007-05-18]. Dostupný z WWW: <http://www.webarchiv.cz/files/dokumenty/seminar/Infoden_061205.ppt>.
- JELÍNKOVÁ, Lenka. 2006. Bibliografický popis elektronických online zdrojů v zahraniční a domácí katalogizační praxi. Praha, 2006. 93 s. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí diplomové práce PhDr. Eva Bratková.
- KIRRIEMUIR, John. 1997. *What is ROADS?* Bath : UKOLN, [October 1997] [cit. 2007-05-18]. Dostupný z WWW: <<http://www.ukoln.ac.uk/metadata/roads/what/>>.
- KOSEK, Jiří. 1999. *XML* [online]. c1999 [cit. 2007-05-18]. Dostupný z WWW: <<http://www.kosek.cz/clanky/xml/xml-uvod.html>>.
- KUČERA, Karel; ŠIBRAVA, František; MAJER, Martin. 2006. *Manuscriptorium v. 1.0. : Manuscriptorium technical compatible = technická kompatibilita* [online]. Verze 1.2. Národní knihovna, 23.2.2006 [cit. 2007-05-18]. Dostupný z WWW: <http://www.manuscriptorium.com/Download/Documentation/manuscriptorium_compatibility_technical_CZE.pdf>.
- LASSILA, Ora. 1997. *Introduction to RDF metadata : W3C note* [online]. W3C, 1997-11-13 [cit. 2007-05-18]. Dostupný z WWW: <<http://www.w3.org/TR/NOTE-rdf-simple-intro-971113.html>>.
- MANOLA, Frank; MILLER, Eric (ed.). 2004. *RDF Primer : W3C Recommendation 10 February 2004* [online]. W3C, 2004 [cit. 2007-08-01]. Dostupný z WWW: <<http://www.w3.org/TR/rdf-primer/>>.
- Masarykova univerzita. 2004. *Dublin Core Czech : Soubor metadatových prvků Dublin Core, verze 1.1 : Referenční popis* [online]. Brno : Masarykova univerzita, 2004-12-20, poslední aktualizace 20. listopad 2006 [cit. 2007-08-09]. Dostupný z WWW: <http://www.ics.muni.cz/dublin_core/elems.html>.
- Masarykova univerzita. 2005. *Dublin Core Czech : Termíny metadat DCMI* [online]. Brno : Masarykova univerzita, 2005-06-13, poslední aktualizace 20. listopad 2006 [cit. 2007-08-09]. Dostupný z WWW: <http://www.ics.muni.cz/dublin_core/terms.html>.
- *MASTER: a gentle introduction* [online]. Centre for Technology and the Arts, c1999, last revised 14 January 2001 [cit. 2007-08-27]. Dostupný z WWW: <<http://xml.coverpages.org/masterGentintr.html>>.
- MATE. 1998. *Project overview* [online]. 12 May, 1998 [cit. 2007-05-18]. Dostupný z WWW: <<http://mate.nis.sdu.dk/about/summary.html>>.
- MATĚJKA, Lukáš. 2006. *Zpřístupnění archivu českého webu*. Brno, 2006. 69 s. Diplomová práce. Masarykova univerzita, Fakulta informatiky. Vedoucí diplomové práce RNDr. Miroslav Bartošek CSc.
- *Metadata Encoding and Transmission Standard : official web site* [online]. Washington, DC : Library of Congress, April 13, 2007 [cit. 2007-05-18]. Dostupný z WWW: <<http://www.loc.gov/standards/mets/>>.
- Mezinárodní federace knihovnických asociací a institucí. 1993. *ISBD (M) : mezinárodní standardní bibliografický popis pro monografie*. Praha : Národní knihovna ČR, 1993. VII, 62 s. ISBD 80-7050-166-9.

- Mezinárodní federace knihovnických asociací a institucí. 1998. *Functional requirements for bibliographic record : final report*. IFLA Study Group on the Functional Requirements for Bibliographic Record. München : Saur, 1998. 136 s. UBCIM Publications - New Series; vol. 19. Dostupný také z WWW: <<http://www.ifla.org/VII/s13/frbr/frbr.htm>>.
- Mezinárodní federace knihovnických asociací a institucí. 2002. *Funkční požadavky na bibliografické záznamy : závěrečná zpráva*. IFLA Study Group on the Functional Requirements for Bibliographic Records; přeložila Ludmila Celbová. Praha : Národní knihovna ČR, 2002. 117 s. Překlad z roku 2001 dostupný také z WWW: <<http://webarchiv.nkp.cz/frbr.pdf>>. ISBN 80-7050-400-5.
- Mezinárodní federace knihovnických asociací a institucí. 2004. *Mapping ISBD elements to FRBR entity attributes and relationship* [online]. 2004-07-28 [cit. 2007-05-16]. Dostupný z WWW: <<http://www.ifla.org/VII/s13/pubs/ISBD-FRBR-mappingFinal.pdf>>.
- Mezinárodní federace knihovnických asociací a institucí. 2005. *Clasification and indexing section : Working group Functional Requirements for Subject Authority Records (FRSAR)* [online]. IFLA, [2005], latest revision 16 January 2006 [cit. 2007-05-16]. Dostupný z WWW: <<http://www.ifla.org/VII/s29/wgfrsar.htm>>.
- Mezinárodní federace knihovnických asociací a institucí. 2007a. *FRBR review group* [online]. IFLA, latest revision March 22, 2007 [cit. 2007-05-16]. Dostupný z WWW: <<http://www.ifla.org/VII/s13/wgfrbr/wgfrbr.htm>>.
- Mezinárodní federace knihovnických asociací a institucí. 2007b. *Functional requirements for authority data : a conceptual model*. IFLA Working Group on the Functional Requirements for Authority Records (FRANAR). 2007-04-01 [cit. 2007-05-16]. Dostupný také z WWW: <<http://www.ifla.org/VII/d4/Franar-Conceptual-M-Draft-e.pdf>>.
- MILLER, Eric. 1997. An introduction to the Resource Description Framework. *D-Lib Magazine* [online]. May 1998, vol 4, no. 5 [cit. 2007-05-18]. Dostupný z WWW: <<http://www.dlib.org/dlib/may98/miller/05miller.html>>. ISSN 1082-9873.
- MIMAS. 2002. *MIMAS Home Page* [online]. Manchester : MIMAS, University of Manchester, 2002-08-29, modified 2006-11-27 [cit. 2007-09-01]. Dostupný z WWW: <<http://www.mimas.ac.uk/>>.
- MIT Libraries. 2004. *Metadata reference guide : a guide to metadata by the Metadata Advisory Group of the MIT Libraries : TEI (Text Encoding Initiative) Metadata* [online]. MIT Libraries, last updated 02/10/04 [cit. 2007-06-19]. Dostupný z WWW: <<http://libraries.mit.edu/guides/subjects/metadata/standards/tei.html>>.
- MORRISON, Alan; POPHAM, Michael; WIKANDER, Karen. 1998. Creating and dokumenting electronic texts : a guide to good practice. *Guides to good practice* [online]. Oxford : Arts and Humanities Data Service, 1988 [cit. 2007-05-18]. Dostupný z WWW: <<http://ota.ahds.ac.uk/documents/creating/>>. ISSN 1463-5194.
- MOTTTL, Jan. 2001. DOBM x XML : další rozvoj AiP Safe. In *Inforum, Praha, Vysoká škola ekonomická, 29.-31.5.2001* [online]. Praha, 2001 [cit. 2007-08-28]. Dostupný také z WWW: <<http://www.inforum.cz/inforum2001/prispevky/motttl.htm>>.
- MUSCIANO, Chuck; KENNEDY, Bill. 2000. HTML a XHTML : kompletní průvodce. Praha : Computer Press, 2000. 1. vyd. xix, 633 s. ISBN 80-7226-407-9.
- Národní knihovna ČR. 2000. *ANL FULL* [online]. Praha : Národní knihovna ČR, c2000-2003 [cit. 2007-05-18]. Dostupný z WWW: <<http://full.nkp.cz>>.
- Národní knihovna ČR. 2005. *O projektu Memoria* [online]. Národní knihovna, [2005] [cit. 2007-05-18]. Dostupný z WWW: <http://www.manuscriptorium.com/Site/CZE/o_projektu.asp>.
- Národní knihovna ČR. 2006a. *Manuscriptorium – základy a kompatibilita* [online]. Verze 1.2. Národní knihovna, 23.2.2006 [cit. 2007-05-18]. Dostupný z WWW: <http://www.manuscriptorium.com/Download/Documentation/manuscriptorium_basics_and_compatibility_CZE.pdf>.

- Národní knihovna ČR. 2006b. *Spolupracující partneři* [online]. Národní knihovna, [2006] [cit. 2007-05-18]. Dostupný z WWW: <<http://www.manuscriptorium.com/Site/CZE/reference.asp>>.
- National Archives of Australia. 2000. *Australian Government Locator Service (AGLS) introduction Service* [online]. Canberra : National Archives of Australia, created 2000-03-30, last modified 2006-05-23 [cit. 2007-06-12]. Dostupný z WWW: <http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html>. Uvedený odkaz byl funkční 12.6.2007, funkční odkaz se nepodařilo dohledat.
- National Archives of Australia. 2002a. *AGLS metadata element set. Part 1: Reference description* [online]. Version 1.3. . Canberra : National Archives of Australia, Office for Government Online, December 2002 [cit. 2007-06-12]. Dostupný z WWW: <http://www.naa.gov.au/recordkeeping/gov_online/agls/metadata_element_set.html>. ISBN 0-642-34491-4. Uvedený odkaz byl funkční 12.6.2007, funkční odkaz se nepodařilo dohledat.
- National Archives of Australia. 2002b. *AGLS metadata element set. Part 2: Usage guide* [online]. Version 1.3. . Canberra : National Archives of Australia, Office for Government Online, December 2002 [cit. 2005-02-15]. Dostupný z WWW: <http://www.naa.gov.au/recordkeeping/gov_online/agls/metadata_element_set.html>. ISBN 0-642-34491-4. Uvedený odkaz byl funkční 15.2.2005, funkční odkaz se nepodařilo dohledat.
- National Archives of Australia. 2002c. *Commonwealth implementation manual : Australian Government Locator Service (AGLS) metadata* [online]. Version 1.0. Canberra : National Archives of Australia, December 2002 [cit. 2007-06-12]. Dostupný z WWW: <http://www.naa.gov.au/recordkeeping/gov_online/agls/cim/AGLS_implement.rtf>. ISBN 0-642-34443-4.
- NDLTD. 2001. *The guide for Electronic Theses and Dissertations* [online]. 12/06/2001 [cit. 2007-06-12]. Dostupný z WWW: <<http://etdguide.org/etdguide.pdf>>.
- NDLTD. 2004a. *History, Description, and Scope* [online]. [c2004] [cit. 2007-06-12]. Dostupný z WWW: <<http://www.ndltd.org/info/description.en.html>>.
- NDLTD. 2004b. *Networked Digital Library of Theses and Dissertations [: NDLTD official members]* [online]. Updated Oct 24, 2004 [cit. 2007-06-12]. Dostupný z WWW: <<http://tennessee.cc.vt.edu/~lming/cgi-bin/ODL/nm-ui/members/index.htm>>.
- NDLTD. 2007a. *Encoding EDT-MS in MARC21* [online]. Last updated 28.2.2007 [cit. 2007-08-30]. Dostupný z WWW: <http://www.ndltd.org/standards/metadata/etdms_and_marc.html>.
- NDLTD. 2007b. *ETD-ms: an interoperability metadata standard for Electronic Theses and Dissertations* [online]. Edited by Anthony Atkins et al. Version 1.00, revision 2. Last updated 02/28/2007 [cit. 2007-06-12]. Dostupný z WWW: <<http://www.ndltd.org/standards/metadata/current.html>>.
- NEDLIB. 1998. *NEDLIB homepage : factsheet* [online]. Den Haag : Koninklijke Bibliotheek, National Library of the Netherlands, 1998-01-01 [cit. 2007-06-13]. Dostupný z WWW: <<http://nedlib.kb.nl/>>.
- *Nordic Metadata Projects* [online]. Last updated 21 February 2000 [cit. 2007-06-13]. Dostupný z WWW: <<http://www.lib.helsinki.fi/meta/index.html>>.
- OGDEN, Bill. 2004. *TIPSTER Annotation Repository* [online]. OK e-VŠKP, c2004 [cit. 2006-08-16]. Dostupný z WWW: <<http://crl.nmsu.edu/twg.annotation/>>. Uvedený odkaz byl funkční 16.8.2006, funkční odkaz se nepodařilo dohledat.
- Open Archives Initiative. 2004. *The Open Archives Initiative Protocol for Metadata Harvesting : Protocol version 2.0 of 2002-06-14* [online]. OAI, 2004/10/12 [cit. 2007-08-31]. Dostupný z WWW: <<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>>.

- Oxford Text Archive. 1998. *Collection policy* [online]. Version 1.1. Oxford : Arts and Humanities Data Service, last updated on March 16, 1998 [cit. 2007-08-31]. Dostupný z WWW: <http://ota.ahds.ac.uk/publications/ID_AHDS-Publications-Collections-Policy.html>.
- Oxford Text Archive. 2006. *The Oxford Text Archive* [online]. Oxford : Arts and Humanities Data Service, last updated 11 August 2006 [cit. 2007-09-05]. Dostupný z WWW: <<http://ota.ahds.ac.uk/index.html>>.
- PALA, Karel. 1996. Informační technologie a korpusová lingvistika (1). *Zpravodaj ÚTV MU* [online]. 1996, roč. 6, č. 3, s. 8-11 [cit. 2007-06-13]. Dostupný také z WWW: <<http://www.ics.muni.cz/bulletin/issues/vol06num03/pala/pala.html>>. ISSN 1212-0901.
- PITTI, Daniel. 1999. Encoded Archival Description : an introduction and overview. *D-lib Magazine* [online]. November 1999, vol. 5, no. 11 [cit. 2007-06-13]. Dostupný z WWW: <<http://www.dlib.org/dlib/november99/11pitti.html>>. ISSN 1082-9873.
- POKORNÝ, Jaroslav. 2005. Směrem k Sémantickému Webu. In: *Moderní databáze 2005 : 20. ročník : sborník z konference, Hotel Amber, Roudnice n. L., 26.-27.5.2005*. Praha : KOMIX, 2005, s. 15-24. Dostupný také z WWW: <<http://www.cs.cas.cz/semweb/download.php?file=05-10-Pokorny&type=pdf>>. ISBN 80-239-4844-X.
- POWELL, Andy; JOHNSTON, Pete. 2003. Guidelines for implementing Dublin Core in XML [online]. DCMI, 2003-04-02 [cit. 2007-08-29]. Dostupný z WWW: <<http://dublincore.org/documents/dc-xml-guidelines/>>.
- POWELL, James; FOX, Edward A. 1998. Multilingual federated searching across heterogeneous collections. *D-lib magazine* [online]. September 1998 vol. 5, iss. 9 [cit. 2002-11-11]. Dostupný z WWW: <<http://www.dlib.org/dlib/september98/powell/09powell.html>>. ISSN 1082-9873.
- PROFFITT, Merrilee. 2004. RedLightGreen : What we've learned since launch. *RLG Focus* [online]. February 2004, iss. 66 [cit. 2007-09-05]. Dostupný také z WWW: <http://www.rlg.org/en/page.php?Page_ID=12661#article2>.
- ProQuest. 2007. *ProQuest Dissertations and Theses Database* [online]. ProQuest, c2007 [cit. 2007-06-13]. Dostupný z WWW: <http://www.proquest.co.uk/products_pq/descriptions/pqdt.shtml>.
- PSOHLAVEC, Stanislav. 2004. Projekt MEMORIA, rukopisy a staré tisky na internetu. *Národní knihovna*. 2004, roč. 15, č. 1, s. 40-43. ISSN 0862-7487.
- RAMSDEN, Chrestopher; BARKAT, Shaheen; KELLEHER, Joh. 2003. *Summative Evaluation of the Archives Hub (2003)* [online]. Tavistock Institute, 2003 [cit. 2007-06-13]. Dostupný z WWW: <<http://www.archiveshub.ac.uk/tavistock2003.pdf>>.
- *RDA : Resource Description and Access* [online]. B.m. : Joint Steering Committee of AACR, last updated 18 May 2007 [cit. 2007-06-13]. Dostupný z WWW: <<http://www.collectionscanada.ca/jsc/rda.html>>.
- ROMARY, Laurent. 2004. An overview of TEI tagging or, Anyone for pizza? In *Wuerzburg TEI Workshop, 8-9 October 2004* [online]. 2004 [cit. 2007-06-19]. Dostupný z WWW: <<http://www.tei-c.org/Talks/2004/Wuerzburg/tei-contents.ppt>>.
- SCHEID, Barbara L.; FOX, Edward A. 2000. *VTLS Partners with Virginia Tech on NDLTD* [online]. July 2000 [cit. 2007-06-19]. Dostupný z WWW: <<http://www.ndltd.org/news/prVTLS.doc>>.
- SHAW, Philip. 2005. *MKSearch* [online]. MKDoc, last modified 2005-11-21 [cit. 2007-06-19]. Dostupný z WWW: <<http://www.mksearch.mkdoc.org/>>.
- SPERBERG-MCQUEEN, C. M.; BURNARD, Lou. 1998. *a gentle introduction to SGML*. Indian TEX users group, 1998-04-16 [cit. 2007-07-23]. Dostupný z WWW: <<http://www.tug.org/applications/pdftex/gentlesgml.pdf>>.
- SPERBERG-MCQUEEN, C. M.; BURNARD, Lou (ed.). 2007. *TEI P5 : Guidelines for Electronic Text Encoding and Interchange* [online]. TEI Consortium, 2007 [cit. 2007-08-27]. Dostupný z WWW: <<http://www.tei-c.org/release/doc/tei-p5-doc/html/index.html>>.

- SUMMERS, Ed. 2004. Building OAI-PMH Harvesters with Net::AOI::Harvester. *Ariadne* [online]. January 2004, issue 38 [cit. 2007-09-05]. Dostupný z WWW: <<http://www.ariadne.ac.uk/issue38/summers/>>. ISSN 1361-3200.
- ŠEDO, Ila. 2004. *Systémy muzejních digitálních informací na WWW*. Praha, 2004. 149 s., 9 s. příloh. Rigorózní práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí diplomové práce PhDr. Eva Bratková.
- VOZKOVÁ, Veronika. 2002. *Aktuální problémy identifikace a popisu webovských informačních zdrojů*. Praha, 2002. v, 152 s., 41 s. Příloh. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí diplomové práce PhDr. Eva Bratková.
- ŠKODOVÁ, Markéta. 2005. *Strategie archivace elektronických online zdrojů a politika jejich výběru do digitálního archivu (se zaměřením na český systém WebArchiv)*. Praha, 2005. 86 s., 16 s. příl. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí diplomové práce PhDr. Eva Bratková.
- Text Encoding Initiative Consortium. 2001. *English poetry full-text database* [online]. TEI Consortium, [2001] [cit. 2007-06-19]. Dostupný z WWW: <<http://www.tei-c.org/Applications/ch01.xml>>.
- Text Encoding Initiative Consortium. 2003. *Getting started with the MASTER XMETAL template* [online]. TEI Consortium, 3.1.2003 [cit. 2007-06-19]. Dostupný z WWW: <<http://www.tei-c.org.uk/Master/Xmetal/tutorial.html>>.
- Text Encoding Initiative Consortium. 2007a. *TEI P5* [online]. Edited by C. M. Sperberg-McQueen and Lou Burnard. TEI Consortium, 2007 [cit. 2007-06-19]. Dostupný z WWW: <<http://www.tei-c.org/release/doc/tei-p5-doc/html/>>.
- Text Encoding Initiative Consortium. 2007b. *What is the TEI Consortium* [online]. TEI Consortium, [2007] [cit. 2007-06-19]. Dostupný z WWW: <<http://www.tei-c.org/Consortium/>>.
- UHLÍŘ, Zdeněk. 1999. Projekt MASTER a standardizace v oblasti zpracování rukopisů. *Národní knihovna*. 1999, roč. 10, č. 3, s. 109-113. ISSN 0862-7487.
- UHLÍŘ, Zdeněk. 2002. Standard MASTER : katalogizace rukopisů v XML. *Národní knihovna*. 2002, roč. 13, č. 2, s. 84-101. ISSN 0862-7487.
- UHLÍŘ, Zdeněk. 2005. Manuscriptorium as virtual environment for the digital history. In *New Information Technologies and the Future of the Humanities*, Praha, 11.2.2005 [online]. Praha : Národní knihovna ČR, 2005 [cit. 2007-08-28]. Dostupný z WWW: <http://www.nkp.cz/pages/page.php3?page=orst_uhlir_manuscriptorium_virtual_environment.htm>.
- UHLÍŘ, Zdeněk. 2006. *Manuscriptorium v. 1.0 : Výběr a popis dokumentů* [online]. Verze 1.2. Národní knihovna, 23.2.2006 [cit. 2007-09-05]. Dostupný z WWW: <http://www.manuscriptorium.com/Download/Documentation/manuscriptorium_document_description_CZE.pdf>.
- UKOLN. 1998. *Dublin Core in RDF : relation - paper magazine, web version and article* [online]. Bath : UKOLN, 1998-10-13 [cit. 2007-06-19]. Dostupný z WWW: <<http://www.ukoln.ac.uk/metadata/resources/rdf/examples/2>>.
- UNESCO. 2005. *UNESCO Thesaurus* [online]. London : University of London Computer Centre, [November 2003] [cit. 2007-06-19]. Dostupný z WWW: <<http://www.ulcc.ac.uk/unesco/>>.
- University of Sheffield. 2001. *PAROLE and SIMPLE* [online]. 1.6.2001 [cit. 2007-06-19]. Dostupný z WWW: <<http://www.dcs.shef.ac.uk/nlp/funded/parole.html>>.
- Vassar College. 2002. *XCES Corpus Encoding Standard for XML : Schemas* [online]. Beta 0.2. Poughkeepsie : Vassar College, last modified 7 May 2002 [cit. 2007-06-19]. Dostupný z WWW: <<http://www.cs.vassar.edu/XCES/schema/>>.

- VOCHOZKA, Josef. 2001. Resource Description Framework. *Zpravodaj ÚVT MU*. 2001, roč. 11, č. 4, s. 10-14. Dostupný také z WWW: <<http://www.ics.muni.cz/to.cs/bulletin/issues/vol11num04/vochozka/vochozka.html>>. ISSN 1212-0901.
- VOJNAR, Martin. 2005. Nové standardy digitálních knihoven pro dlouhodobou ochranu. *Knihovna* [online]. 2005, roč. 16, č. 2, s. 45-58. Dostupný také z WWW: <<http://knihovna.nkp.cz/knihovna61/vojnar.htm>>. ISSN 1801-3252.
- VOJTÁŠEK, Filip. 2001. Projekty zaměřené na archivaci a zpřístupnění webových zdrojů. In *Automatizace knihoven*, 25.9.2001 [online]. 2001 [cit. 2007-09-05]. Dostupný také z WWW: <http://www.webarchiv.cz/files/dokumenty/konference/aut2001_fv.zip>.
- VOJTÁŠEK, Filip. 2002. *RE: WebArchiv* [elektronická pošta]. Message to: Veronika Vozkova. 2002-11-12 16:06 [cit. 2002-12-12]. Osobní komunikace.
- VOZKOVÁ, Veronika. 2001a. *Problematika identifikace a popisu webovských informačních zdrojů : analýza vybraných současných řešení : bibliografická rešerše*. Praha, 2001. 24 s. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví.
- VOZKOVÁ, Veronika. 2001b. *Systém Dublin Core a jeho metadata : studijně rozborová práce*. Praha, 2001. 32 s. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví.
- VOZKOVÁ, Veronika. 2002. *Aktuální problémy identifikace a popisu webovských informačních zdrojů*. Praha, 2002. v, 152 s., 41 s. Příloh. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí diplomové práce PhDr. Eva Bratková.
- VTLS. 2004. *Vitrua is designed for libraries that expect more than meet the eyes* [online]. Blacksburg (VA) : VTLS, [2004] [cit. 2007-06-19]. Dostupný z WWW: <<http://www.vtls.com/brochures/virtua.pdf>>.
- WHITE, Gerry; IVANOVA, Nelly. 2003. Education Network Australia : Discover, communicate, collaborate!. In *British Council Meeting, February 2003* [online]. Education.au, 2003 [cit. 2007-08-23]. Dostupný z WWW: <http://www.educationau.edu.au/jahia/webdav/site/myjahiasite/shared/papers/BC_paper_final.pdf>.
- WHITE, Gerry. 2005. *Australian education content building : building Online learning in a borderless world* [online]. Education.au, last modified 02 Nov 2005 [cit. 2007-07-23]. Dostupný z WWW: <<http://www.educationau.edu.au/jahia/webdav/site/myjahiasite/shared/papers/berlin.pdf>>.
- WILSON, Andrew; WARD, Nigel. 2002a. *AGLS controlled vocabularies : audience controlled vocabulary* [online]. Canberra : National Archives of Australia, 2002-02-19 [cit. 2007-06-19]. Dostupný z WWW: <http://www.naa.gov.au/recordkeeping/gov_online/agls/schemes/agls-audience1.0.html>.
- WILSON, Andrew; WARD, Nigel. 2002b. *AGLS encoding schemes : agent encoding scheme* [online]. Canberra : National Archives of Australia, 2002-02-19 [cit. 2007-06-19]. Dostupný z WWW: <http://www.naa.gov.au/recordkeeping/gov_online/agls/schemes/AglsAgent1.0.html>.
- WILSON, Andrew; WARD, Nigel. 2002c. *AGLS encoding schemes : availability encoding scheme* [online]. Canberra : National Archives of Australia, 2002-02-19 [cit. 2007-06-19]. Dostupný z WWW: <http://www.naa.gov.au/recordkeeping/gov_online/agls/schemes/AglsAvail1.0.html>.
- WILSON, Andrew; WARD, Nigel. 2002d. *AGLS encoding schemes : document type controlled vocabulary* [online]. Canberra : National Archives of Australia, 2002-02-19 [cit. 2007-06-19]. Dostupný z WWW: <http://www.naa.gov.au/recordkeeping/gov_online/agls/schemes/agls-document1.0.html>.

- WILSON, Andrew; WARD, Nigel. 2002e. *AGLS encoding schemes : jurisdiction controlled vocabulary* [online]. Canberra : National Archives of Australia, 2002-02-19 [cit. 2007-06-19]. Dostupný z WWW: <http://www.naa.gov.au/recordkeeping/gov_online/agls/schemes/AglsJuri1.0.html>.
- WILSON, Andrew; WARD, Nigel. 2002f. *AGLS encoding schemes : service type controlled vocabulary* [online]. Canberra : National Archives of Australia, 2002-02-19 [cit. 2007-06-19]. Dostupný z WWW: <http://www.naa.gov.au/recordkeeping/gov_online/agls/schemes/agls-service1.0.html>.
- WUEPPELMANN, William. 2006. *RE: question – metadata in ECO* [elektronická pošta]. Message to: Veronika Synková, 2006-12-08 19:04 [cit. 2006-12-10]. Osobní komunikace.
- ŽABIČKA, Petr. 2000. NEDLIB Harvester - technika "sklizené" informací. *Ikaros* [online]. 2000, roč. 4, č. 10 [cit. 2007-06-19]. Dostupný na World Wide Web: <<http://www.ikaros.cz/node/672>>. ISSN 1212-5075.
- ŽABIČKA, Petr. 2002a. Infrastruktura Webarchivu v roce 2002. In *Inforum. Sekce: Digitalizace sbírek a digitální přístup k dokumentům, Praha, Vysoká škola ekonomická, 23.5.2002* [online]. Praha, 2002 [cit. 2007-06-19]. Dostupný z WWW: <<http://www.inforum.cz/inforum2002/prednaska8.htm>>.
- ŽABIČKA, Petr. 2002b. *Re: WebArchiv* [elektronická pošta]. Message to: Veronika Vozkova. 2002-10-31 17:09 [cit. 2002-11-11]. Osobní komunikace.
- ŽABIČKA, Petr. 2005. Infrastruktura projektu WebArchiv. In *Informační setkání u příležitosti zpřístupnění digitálního archivu, Národní knihovna ČR, 6.12.2005* [online]. WebArchiv, 2005 [cit. 2007-09-05]. Dostupný z WWW: <http://www.webarchiv.cz/files/dokumenty/seminar/Infoden_061205.ppt>.

PŘÍLOHY

Příloha 1: Kvalifikovaný záznam Dublin Core ve formátu XML

```
<?xml version="1.0"?>
<metadata
  xmlns="http://example.org/myapp/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://example.org/myapp/ http://example.org/myapp/schema.xsd"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/">
  <dc:title>
    UKOLN
  </dc:title>
  <dcterms:alternative>
    UK Office for Library and Information Networking
  </dcterms:alternative>
  <dc:subject>
    national centre, network information support, library community, awareness, research, information
    services,public library networking, bibliographic management, distributed library systems,
    metadata, resource discovery, conferences,lectures, workshops
  </dc:subject>
  <dc:subject xsi:type="dcterms:DDC">
    062
  </dc:subject>
  <dc:subject xsi:type="dcterms:UDC">
    061(410)
  </dc:subject>
  <dc:description>
    UKOLN is a national focus of expertise in digital information management. It provides policy,
    research and awareness services to the UK library, information and cultural heritage communities.
    UKOLN is based at the University of Bath.
  </dc:description>
  <dc:description xml:lang="fr">
    UKOLN est un centre national d'expertise dans la gestion de l'information digitale.
  </dc:description>
  <dc:publisher>
    UKOLN, University of Bath
  </dc:publisher>
  <dcterms:isPartOf xsi:type="dcterms:URI">
    http://www.bath.ac.uk/
  </dcterms:isPartOf>
  <dc:identifier xsi:type="dcterms:URI">
    http://www.ukoln.ac.uk/
  </dc:identifier>
  <dcterms:modified xsi:type="dcterms:W3CDTF">
    2001-07-18
  </dcterms:modified>
  <dc:format xsi:type="dcterms:IMT">
    text/html
  </dc:format>
  <dcterms:extent>
    14 Kbytes
  </dcterms:extent>
</metadata>
```

(Powell, 2003)

Příloha 2: Prvky formátu EdNA

Prvek	Kvalifikátor	Popis	Schéma
DC.Identifier		Jedinečný identifikátor zdroje	URI
DC.Title		Jméno zdroje	
	Alternative	Jakákoliv jiná forma jména zdroje	
DC.Description		Popis obsahu zdroje	
	Table of Contents	Seznam částí zdroje	
	Abstract	Abstrakt, stručný popis zdroje	
DC.Subject		Téma zdroje	ASCED APSDEP DDC edna-kla LCC LCSH MeSH SCIS UDC edna-kla: values (arts, english, english literacy, enterprise, health and physical education, languages other than English, mathematics, numeracy, science, studies of society and environment, technology, vocational learning)
DC.Publisher		Entita odpovědná za zveřejnění a dostupnost zdroje.	
DC.Creator		Entita primárně odpovědná za obsah zdroje.	
DC.Date		Datum spojené s životním cyklem zdroje.	DCMI Period W3C-DTF
	Created	Datum vytvoření zdroje.	
	Valid	Datum (nebo časové rozmezí) platnosti zdroje	
	Available	Datum (nebo časové rozmezí) dostupnosti zdroje	
	Issued	Datum formálního zveřejnění zdroje.	
	Modified	Datum změny obsahu zdroje.	

Prvek	Kvalifikátor	Popis	Schéma
DC.Type		Žánr obsahu zdroje	DCMI Type Vocabulary (collection, dataset, event, image, interactive resource, service, software, sound, text). edna-document (bibliography, dissertation, guidelines, index, manuscript, policy document, presentation, reference, report, research report, serial) edna-curriculum (activity, assessment, course, curriculum/syllabus, exemplar, lesson plan, online project, training package, unit/module). edna-event (conference, exhibition, performance).
DC.Format		Fyzické nebo digitální vyjádření zdroje	
	Extent	Velikost zdroje	
	Medium	Materiál nebo nosič zdroje	
DC.Language		Jazyk zdroje	ISO 639-2 RFC 1766
DC.Coverage	Spatial	Prostorové charakteristiky zdroje	edna-spatial ACT - Australian Capital Territory NSW - New South Wales NT - Northern Territory QLD - Queensland SA - South Australia TAS - Tasmania VIC - Victoria WA - Western Australia Australia DCMI Point ISO 3166 DCMI Box TGN
DC.Rights		Informace o autorských právech vztahujících se ke zdroji	

Prvek	Kvalifikátor	Popis	Schéma
DC.Relation		Vyjádření vztahu k souvisejícímu zdroji	URI
	Is Version Of	Popisovaný zdroj je verzí odkazovaného zdroje	
	Has Version	Odkazovaný zdroj je verzí popisovaného zdroje	
	Is Replaced By	Popisovaný zdroj byl nahrazen odkazovaným zdrojem	
	Is Required By	Z fyzického nebo logického hlediska je popisovaný zdroj potřebný k interpretaci odkazovaného zdroje	
	Requires	Popisovaný zdroj potřebuje z fyzického nebo logického hlediska k interpretaci odkazovaný zdroj	
	Is Part Of	Popisovaný zdroj je fyzickou nebo logickou částí odkazovaného zdroje	
	Has Part	Odkazovaný zdroj je fyzickou nebo logickou částí popisovaného zdroje	
	Is Referenced By	Odkazovaný zdroj cituje popisovaný zdroj	
	References	Popisovaný zdroj cituje odkazovaný zdroj	
	Is Format Of	Popisovaný zdroj nese stejný obsah v jiném formátu než odkazovaný zdroj. Prvotní byl odkazovaný zdroj	
	Has Format	Odkazovaný zdroj byl vytvořen se stejným obsahem jako zdroj popisovaný, ale má jiný obsah. Prvotní byl popisovaný zdroj	
DC.Contributor		Entita odpovědná za určitou část obsahu zdroje	
DC.Source		Odkaz na zdroj, ze kterého byl popisovaný zdroj odvozen	URI

(EdNA, 2002)

Příloha 3: Konverzní tabulka pro formáty ETD-MS a MARC21

ETD	MARC21
dc.title	245a
dc.title.alternative	246
dc.title.translated	242
dc.creator	100a
dc.subject	653a
dc.description.abstract	520a
dc.description.note	Jakékoliv nepoužité pole 5xx, např. 504
dc.description.release	
dc.publisher	260a+b
dc.contributor	720a
dc.contributor.role	720e
dc.date	008 pozice 7-10
dc.type	hlavička 6&7; 655
dc.format	856q
dc.identifier	856u
dc.language	008 pozice 35-37; [546]
dc.coverage	651 nebo 690
dc.rights	540
thesis.degree.name	502a
thesis.degree.level	502a
thesis.degree.discipline	710b
thesis.degree.grantor	502a; 710ab

(NDLTD, 2007a)

Příloha 4: Prvky používané v databázi WebArchivu

Prvek	Popis	Kvalifikátor nebo schéma
Název	Jméno dané zdroji	Název Alternativní název
Tvůrce	Entita primárně zodpovědná za vytvoření obsahu zdroje	Tvůrce Jméno osoby: „Příjmení, rodné jméno“ Jméno korporace (organizace) (E-mailová) adresa tvůrce
Předmět: klíčová slova	Vlastní klíčová slova	-
Předmět: řízené slovníky	Klíčová slova z řízených slovníků	PHNK – Předmětová hesla Národní knihovny PSH – Polytematický strukturovaný heslář LCSH - Library of Congress Subject Headings MeSH – Medical Subject Headings EUROVOC ČTT – Český teologický tezaurus AGROVOC
Předmět: třídění	Třídění	MDT- Mezinárodní desetinné třídění MDT_KON - Konspekt DDT – Deweyho Desetinné třídění DDT_KON - Konspekt LCC - Třídění Kongresové knihovny
Popis	Vysvětlení obsahu zdroje	Anotace Obsah
Vydavatel	Entita odpovědná za zpřístupnění zdroje	Vydavatel (E-mailová) adresa vydavatele
Přispěvatel	Entita, která přispěla k vytvoření obsahu zdroje	Přispěvatel Jméno osoby: „Příjmení, rodné jméno“ Jméno korporace (organizace) Emailová adresa přispěvatele
Datum	Datum spojené s určitou událostí během existence zdroje Je možné definovat jak konkrétní datum, tak rozmezí dat	Nedefinováno Vytvořeno Vydáno Platnost Zpřístupněno Změněno
Typ: DCMI	Povaha nebo druh obsahu zdroje	Text Pohyblivý obraz Fyzický objekt Statický obraz Obraz Zvuk Software Interaktivní zdroj Služba Událost Databáze Sbírka
Typ: PHNK	Typ, povaha, žánr dokumentu	PHNK (Předmětová hesla Národní knihovny ČR)

Prvek	Popis	Kvalifikátor nebo schéma	
Formát: formát souboru	Formát (souboru) zdroje	Nedefinováno text/html text/xml (.xml) text/plain (.txt) aplikace/zip archiv (.zip) aplikace/ms word (.doc) aplikace/ms excel (.xls) aplikace/pdf (.pdf) aplikace/postscript (.ai, .eps, .ps) aplikace/ms powerpoint (.ppt) aplikace/rtf (.rtf) aplikace/x-latex (.latex) obraz/gif (.gif) obraz/jpeg (.jpeg, .jpe, .jpg) audio/mpeg video/mpeg (.mpg, .mpeg)	
Formát: médium	Fyzický nosič zdroje	Nedefinováno Počítačový soubor CD-ROM DVD-disk	
Formát: rozsah	Velikost nebo rozsah zdroje	-	
Identifikátor: URL	URL-adresa zdroje	-	
Identifikátor	Znakový řetězec nebo číslo, které jednoznačně identifikuje zdroj	URN ISBN ISSN SICI	
Zdroj	Odkaz na zdroj, z něhož je popisovaný zdroj odvozen	URL URN ISBN ISSN SICI Vlný text	
Jazyk	Jazyk intelektuálního obsahu zdroje	Angličtina Francouzština Němčina Italština Španělština Čeština Slovenština Nedefinováno	
Vztah	Odkaz na příbuzný zdroj	Je verzí Má verzi Nahrazen Nahrazuje Je požadovaný Požaduje Je částí Obsahuje část Je odkazovaný Odkazuje Je formátem Má formát	URL URN ISBN Text

Prvek	Popis	Kvalifikátor nebo schéma	
Pokrytí	Rozsah nebo záběr obsahu zdroje	Prostorové Časové Nedefinováno	PHNK (geografické) DCMI Point (prostor) ISO3166 (prostor) DCMI Box (prostor) TGN (prostor) DCMI Period (čas) ISO 8601 (čas) Volný text
Práva	Informace o právech vztahujících se k popisovanému zdroji	-	

(Generátor metadat, 2001)

Příloha 5: Kvalifikovaný záznam vyhledaný v databázi ANL FULL

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcq="http://purl.org/dc/documents/rec/dcmes-
qualifiers-20000711.htm">
  <rdf:Description rdf:about="Koukněte, jak žijeme">
    <dc:Title>Koukněte, jak žijeme</dc:Title>
    <dc:Subtitle>Film o rodině, která se nechce lišit</dc:Subtitle>
    <dc:Creator>
      <rdf:Description>
        <dcq:CreatorType>personalName</dcq:CreatorType>
        <rdf:value>Mandys, Pavel</rdf:value>
      </rdf:Description>
    </dc:Creator>
    <dc:Subject>
      <rdf:Bag>
        <rdf:li>čeští režiséři</rdf:li>
      </rdf:li>
      <rdf:Description>
        <dcq:SubjectScheme>ANL-KATEGORIE</dcq:SubjectScheme>
        <rdf:value>kinematografie: česká kinematografie</rdf:value>
      </rdf:Description>
      </rdf:li>
      <rdf:li>
        <rdf:Description>
          <dcq:SubjectScheme>ANL-KATEGORIE</dcq:SubjectScheme>
          <rdf:value>filmy: české hrané filmy</rdf:value>
        </rdf:Description>
      </rdf:li>
      <rdf:li>
        <rdf:Description>
          <dcq:SubjectScheme>MDT</dcq:SubjectScheme>
          <rdf:value>791.43</rdf:value>
        </rdf:Description>
      </rdf:li>
      <rdf:li>
        <rdf:Description>
          <dcq:SubjectScheme>MDT</dcq:SubjectScheme>
          <rdf:value>791.44.071.1</rdf:value>
        </rdf:Description>
      </rdf:li>
      <rdf:li>
        <rdf:Description>
          <dcq:SubjectScheme>MDT</dcq:SubjectScheme>
          <rdf:value>(437.3)</rdf:value>
        </rdf:Description>
      </rdf:li>
      <rdf:li>
        <rdf:Description>
          <dcq:SubjectScheme>PHNK</dcq:SubjectScheme>
          <rdf:value>Tyc, Zdeněk</rdf:value>
        </rdf:Description>
      </rdf:li>
    </dc:Subject>
  </rdf:Description>
</rdf:RDF>
```

```

- <rdf:li>
- <rdf:Description>
-   <dcq:SubjectScheme>PHNK</dcq:SubjectScheme>
-   <rdf:value>Smradi (film)</rdf:value>
- </rdf:Description>
- </rdf:li>
- <rdf:li>
- <rdf:Description>
-   <dcq:SubjectScheme>PHNK</dcq:SubjectScheme>
-   <rdf:value>hrané filmy-Česko-r. 2002</rdf:value>
- </rdf:Description>
- </rdf:li>
- <rdf:li>
- <rdf:Description>
-   <dcq:SubjectScheme>PHNK</dcq:SubjectScheme>
-   <rdf:value>režiséři-Česko-r. 2002</rdf:value>
- </rdf:Description>
- </rdf:li>
- </rdf:Bag>
- </dc:Subject>
- <dc:Coverage>
- <rdf:Bag>
- <rdf:li>
- <rdf:Description>
-   <dcq:CoverageType>spatial</dcq:CoverageType>
-   <rdf:value>Česko</rdf:value>
- </rdf:Description>
- </rdf:li>
- <rdf:li>
- <rdf:Description>
-   <dcq:CoverageType>temporal</dcq:CoverageType>
-   <rdf:value>r. 2002</rdf:value>
- </rdf:Description>
- </rdf:li>
- </rdf:li>
- <rdf:Description>
-   <dcq:CoverageType>spatial</dcq:CoverageType>
-   <rdf:value>Česko</rdf:value>
- </rdf:Description>
- </rdf:li>
- <rdf:li>
- <rdf:Description>
-   <dcq:CoverageType>temporal</dcq:CoverageType>
-   <rdf:value>r. 2002</rdf:value>
- </rdf:Description>
- </rdf:li>
- </rdf:li>
- <rdf:Description>
-   <dcq:CoverageScheme>GAC</dcq:CoverageScheme>
-   <dcq:CoverageType>spatial</dcq:CoverageType>
-   <rdf:value>e-xr---</rdf:value>
- </rdf:Description>
- </rdf:li>
- </rdf:Bag>
- </dc:Coverage>

```

```

<dc:Description>
<rdf:Description>
  <dcq:DescriptionType>abstract</dcq:DescriptionType>
  <rdf:value>Tereza Boučková přinesla scénář, režiséru Zdeňku Tycovi se zalíbil a producent
Vratislav Šlajer na to sehnal peníze</rdf:value>
</rdf:Description>
</dc:Description>
<dc:Publisher>Mediacop</dc:Publisher>
<dc>Date>
<rdf:Bag>
<rdf:li>
<rdf:Description>
  <dcq:DateScheme>W3C-DTF</dcq:DateScheme>
  <dcq:DateType>issued</dcq:DateType>
  <rdf:value>2002-09-30</rdf:value>
</rdf:Description>
</rdf:li>
<rdf:li>
<rdf:Description>
  <dcq:DateScheme>W3C-DTF</dcq:DateScheme>
  <dcq:DateType>modified</dcq:DateType>
  <rdf:value>2002-10-23</rdf:value>
</rdf:Description>
</rdf:li>
</rdf:Bag>
</dc>Date>
<dc>Type>
<rdf:Bag>
<rdf:li>
<rdf:Description>
  <dcq:TypeScheme>DCMI</dcq:TypeScheme>
  <rdf:value>text</rdf:value>
</rdf:Description>
</rdf:li>
<rdf:li>article</rdf:li>
<rdf:li>článek</rdf:li>
<rdf:li>recenze</rdf:li>
</rdf:Bag>
</dc>Type>
<dc:Format>
<rdf:Description>
  <dcq:FormatScheme>IMT</dcq:FormatScheme>
  <rdf:value>text/html</rdf:value>
</rdf:Description>
</dc:Format>
<dc:Identifier>
<rdf:Bag>
<rdf:li>http://full.nkp.cz/nkdb/view/mkdoc.asp?p1=TY2002093000007500</rdf:li>
<rdf:li>
<rdf:Description>
  <dcq:IdentifierScheme>SICI</dcq:IdentifierScheme>
  <rdf:value>1210-9940(20020930)9:40<81:TY2002093000007500>3.0.CO;2-X</rdf:value>
</rdf:Description>
</rdf:li>

```

```

=> <rdf:li>
=> <rdf:Description>
  <dcq:IdentifierScheme>URN</dcq:IdentifierScheme>
  <rdf:value>URN:NBN:cz-nkTY2002093000007500</rdf:value>
</rdf:Description>
</rdf:li>
</rdf:Bag>
</dc:Identifier>
=> <dc:Source>
=> <rdf:Bag>
  <rdf:li>Týden roč. 9, č. 40 (2002-09-30), s. 81</rdf:li>
=> <rdf:li>
=> <rdf:Description>
  <dcq:SourceScheme>ISSN</dcq:SourceScheme>
  <rdf:value>1210-9940</rdf:value>
</rdf:Description>
</rdf:li>
</rdf:Bag>
</dc:Source>
=> <dc:Language>
=> <rdf:Description>
  <dcq:LanguageScheme>RFC1766</dcq:LanguageScheme>
  <rdf:value>cs</rdf:value>
</rdf:Description>
</dc:Language>
=> <dc:Rights>
=> <rdf:Bag>
  <rdf:li>Mediacop</rdf:li>
  <rdf:li>Anopress</rdf:li>
</rdf:Bag>
</dc:Rights>
<dc:Text>Plný text článku pouze pro registrované uživatele...</dc:Text>
</rdf:Description>
</rdf:RDF>

```

Příloha 6: Ukázka hlavičky TEI

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Common sense, a machine-readable transcript</title>
      <author>Paine, Thomas (1737-1809)</author>
      <respStmt>
        <resp>compiled by</resp>
        <name>Jon K Adams</name>
      </respStmt>
    </titleStmt>
    <editionStmt>
      <edition>
        <date>1986</date>
      </edition>
    </editionStmt>
    <publicationStmt>
      <distributor>Oxford Text Archive.</distributor>
      <address>
        <addrLine>Oxford University Computing Services,</addrLine>
        <addrLine>13 Banbury Road,</addrLine>
        <addrLine>Oxford OX2 6RB,</addrLine>
        <addrLine>UK</addrLine>
      </address>
    </publicationStmt>
    <notesStmt>
      <note>Brief notes on the text are in a supplementary file.</note>
    </notesStmt>
    <sourceDesc>
      <biblStruct>
        <monogr>
          <editor>Foner, Philip S.</editor>
          <title>The collected writings of Thomas Paine</title>
          <imprint>
            <pubPlace>New York</pubPlace>
            <publisher>Citadel Press</publisher>
            <date>1945</date>
          </imprint>
        </monogr>
      </biblStruct>
    </sourceDesc>
  </fileDesc>
  <encodingDesc>
    <samplingDecl>
      <p>Editorial notes in the Foner edition have not been reproduced. </p>
      <p>Blank lines and multiple blank spaces, including paragraph indents, have not been preserved.
    </p>
    </samplingDecl>
    <editorialDecl>
      <correction status="high" method="silent">
        <p>The following errors in the Foner edition have been corrected:
        <list>
          <item>p. 13 l. 7 cotemporaries contemporaries </item>
          <item>p. 28 l. 26 [comma] [period] </item>
          <item>p. 84 l. 4 kin kind </item>
          <item>p. 101 l. 4 certainy certainty </item>
        </list>
        </p>
      </correction>
    </editorialDecl>
  </encodingDesc>
</teiHeader>
```

```

</correction>
<normalization>
  <p>No normalization beyond that performed by Foner, if any. </p>
</normalization>
<quotation marks="all" form="std">
  <p>All double quotation marks rendered with ", all single quotation marks with  apostrophe.
</p>
</quotation>
<hyphenation eol="none">
  <p>Hyphenated words that appear at the end of the line in the Foner edition have been
reformed.</p>
</hyphenation>
<stdVals>
  <p>Standard date values are given in ISO form: yyyy-mm-dd. </p>
</stdVals>
<interpretation>
  <p>Compound proper names are marked. </p>
  <p>Dates are marked. </p>
  <p>Italics are recorded without interpretation. </p>
</interpretation>
</editorialDecl>
<classDecl>
  <taxonomy xml:id="lcsh">
    <bibl>Library of Congress Subject Headings</bibl>
  </taxonomy>
  <taxonomy xml:id="lc">
    <bibl>Library of Congress Classification</bibl>
  </taxonomy>
</classDecl>
</encodingDesc>
<profileDesc>
  <creation>
    <date>1774</date>
  </creation>
  <langUsage>
    <language ident="en" usage="100">English.</language>
  </langUsage>
  <textClass>
    <keywords scheme="#lcsh">
      <list>
        <item>Political science</item>
        <item>United States -- Politics and government — Revolution, 1775-1783</item>
      </list>
    </keywords>
    <classCode scheme="#lc">JC 177</classCode>
  </textClass>
</profileDesc>
<revisionDesc>
  <change date="1995-07-04">
    <name>R.G. </name> finished data entry
  </change>
  <change date="1995-01-15">
    <name>R.G. </name> began data entry
  </change>
</revisionDesc>
</teiHeader>

```

(Text Encoding Initiative Consortium, 2007a)

Příloha 7: Hlavička TEI k záznamu z Oxfordského textového archivu

```
<!DOCTYPE tei.2 PUBLIC "-//TEI//DTD TEI Lite 1.0//EN">
<TEI.2>
<TEIHEADER TYPE="ISBD(ER)" DATE.CREATED="1997-12-18">
<FILEDESC>
<TITLESTMT>
<TITLE TYPE="main">Jane Eyre [Electronic resource] / by Charlotte Bronte [sic]</TITLE>
<AUTHOR>Brontë, Charlotte, 1816-1855</AUTHOR>
<RESPSTMT>
<RESP>creation of machine-readable version</RESP>
<NAME>Project Eris</NAME>
</RESPSTMT>
</TITLESTMT>
<EXTENT>
<SEG TYPE="designation">Text data</SEG>
<SEG TYPE="size">(1 file : ca. 0.98 megabytes)</SEG>
</EXTENT>
<PUBLICATIONSTMT>
<authority>
<!-- Confidential Information Removed -->
<date>1993-12-21</date>
</authority>
<DISTRIBUTOR>
<NAME KEY="ota" TYPE="organisation">Oxford Text Archive</NAME>
<NAME TYPE="place">Oxford</NAME>
<ADDRESS>
<ADDRLINE>
<NAME KEY="oucs" TYPE="organisation">Oxford University Computing Services</NAME>
</ADDRLINE>
<ADDRLINE>13 Banbury Road</ADDRLINE>
<ADDRLINE>Oxford</ADDRLINE>
<ADDRLINE>OX2 6NN</ADDRLINE>
<ADDRLINE>
<NAME TYPE="email">info@ota.ahds.ac.uk</NAME>
</ADDRLINE>
</ADDRESS>
</DISTRIBUTOR>
<IDNO TYPE="ota">bron2001</IDNO>
<AVAILABILITY STATUS="FREE">
<P>Freely available for non-commercial use provided that this header is included in its entirety with
any copy distributed</P>
</AVAILABILITY>
</PUBLICATIONSTMT>
<NOTESSTMT>
<NOTE>Mode of access: Online. OTA website</NOTE>
<NOTE>Title proper taken from title in electronic text</NOTE>
</NOTESSTMT>
<SOURCEDESC>
<BIBL>
<NOTE>The source is claimed to be an ed. printed in 1846 which must be the 1st ed. but no other
details are given for clarification.</NOTE>
</BIBL>
</SOURCEDESC>
</FILEDESC>
```

```

<ENCODINGDESC>
<PROJECTDESC>
<P>Project Eris is a major gopher-based collection of world classics in English, compiled by Virginia
Tech, but now defunct at that website</P>
</PROJECTDESC>
<SAMPLINGDECL>
<P>
</P>
</SAMPLINGDECL>
<EDITORIALDECL>
<P>Encoding format: Plain text</P>
</EDITORIALDECL>
<TAGSDECL>
</TAGSDECL>
<REFSDECL>
<P>
</P>
</REFSDECL>
<CLASSDECL>
<TAXONOMY ID="OTASH">
<BIBL>Oxford Text Archive Subject Headings</BIBL>
</TAXONOMY>
<TAXONOMY ID="LCSH">
<BIBL>Library of Congress Subject Headings</BIBL>
</TAXONOMY>
</CLASSDECL>
</ENCODINGDESC>
<PROFILEDESC>
<CREATION>
<DATE>
</DATE>
</CREATION>
<LANGUSAGE>
<LANGUAGE ID="eng">English</LANGUAGE>
</LANGUSAGE>
<TEXTCLASS>
<KEYWORDS SCHEME="OTASH">
<TERM TYPE="genre">
</TERM>
</KEYWORDS>
<KEYWORDS SCHEME="LCSH">
<TERM TYPE="genre">Novels -- Great Britain -- 19th century</TERM>
</KEYWORDS>
</TEXTCLASS>
</PROFILEDESC>
<REVISIONDESC>
<CHANGE>
<DATE>2000-09-29</DATE>
<RESPSTMT>
<RESP>Cataloguer</RESP>
<NAME>Webb, Anton</NAME>
</RESPSTMT>
<ITEM>Header changed in accordance with ISBD(ER) guidelines and expanded. Validated in the
TEI Lite DTD using XMetal 2.0</ITEM>
</CHANGE>
</REVISIONDESC>
</TEIHEADER>
</TEI.2>

```

Příloha 8: Hlavička CES

```
<cesHeader version="2.0">
  <fileDesc>
    <titleStmt>
      <h.title>Machine-readable version of 1984, ch. 1</h.title>
    </titleStmt>
    <extent>
      <wordcount>6571 </wordcount>
      <bytecount units="bytes">6571 </bytecount>
    </extent>
    <publicationStmt>
      <distributor>Laboratoire Parole et Langage, CNRS</distributor>
      <pubAddress>29, avenue Robert Schuman Aix-en-Provence, France</pubAddress>
      <telephone>+33 42 95 36 33</telephone>
      <fax>+33 42 59 50 96</fax>
      <eAddress>phonetic@univ-aix.fr</eAddress>
      <availability status="restricted">internal use only--cannot be distributed</availability>
      <pubDate>6571</pubDate>
    </publicationStmt>
    <sourceDesc>
      <biblStruct>
        <monogr>
          <h.title>Nineteen Eighty-four</h.title>
          <h.author>George Orwell</h.author>
          <imprint>
            <pubPlace>New York</pubPlace>
            <publisher>New American Library</publisher>
            <pubDate>1949; reprinted 1961</pubDate>
          </imprint>
        </monogr>
      </biblStruct>
    </sourceDesc>
  </fileDesc>
  <encodingdesc>
    <projectdesc>
      This English version of the first chapter of Orwell's 1984 is encoded for use in the
      MULTEXT-EAST project. The English is to serve as the base for the parallel corpus, and will
      be aligned to versions of the text in Romanian, Bulgarian, Estonian, Slovenian, Czech,
      and Hungarian.
    </projectdesc>
    <editorialdecl>
      <correction status="medium method=silent"></correction>
      <segmentation>Marked up to the level of paragraph plus marking of particular sub-
      paragraph elements: NAME, DATE, FOREIGN.
    </segmentation>
    </editorialdecl>
    <tagsdecl>
      <tagusage gi="body" occurs="1"></tagusage>
      <tagusage gi="date" occurs="5"></tagusage>
      <tagusage gi="div" occurs="2"></tagusage>
      <tagusage gi="foreign" occurs="4"></tagusage>
      <tagusage gi="hi" occurs="4"></tagusage>
      <tagusage gi="name" occurs="149"></tagusage>
      <tagusage gi="note" occurs="1"></tagusage>
      <tagusage gi="num" occurs="2"></tagusage>
      <tagusage gi="quote" occurs="3"></tagusage>
    </tagsdecl>
  </encodingdesc>
```

```

<profiledesc>
  <language>
    <language id="fr" iso639="fr">French</language>
    <language id="en" iso639="en">English</language>
    <language id="la" iso639="la">Latin</language>
    <language id="ns">Newspeak</language>
  </language>
  <wsdUsage>
    <writingSystem id="ISO 8859-1">ISO character set for western European
languages</writingSystem>
    <writingSystem id="ISO 8859-5">ISO character set for Cyrillic</writingSystem>
  </wsdUsage>
</profiledesc>
</cesHeader>

```

(Ide, 1996)

Příloha 9: Prvky formátu EAD

Prvek	Popis
<abbr> Abbreviation	prvek určený pro zápis zkratk a akronymů
<abstract> Abstract	stručné shrnutí textu
<accessrestrict> Restrictions on Access	omezení přístupu ke zdroji (právní status, archiv, apod.)
<accruals> Accruals	podřízený prvek prvku <admininfo> indikující kvantitu či frekvenci předpokládaných změn zdroje
<acqinfo> Acquisition Information	podřízený prvek prvku <admininfo> pro informace o zdroji a způsobu akvizice
<address> Address	informace o adrese
<addressline> Address Line	jedna řádka adresy
<altformavail> Alternative Form Available	podřízený prvek prvku <admininfo> zpřístupňující informace o dalších exemplářích popisovaného zdroje
<appraisal> Appraisal Information	podřízený prvek prvku <admininfo>, obsahuje údaje o informační hodnotě zdroje, jeho stavu, využitelnosti a vztahu k jiným zdrojům
<arc> Arc	slouží pro definici odkazů mezi pojmenovanými páry zdrojů
<archdesc> Archival Description	popisuje kontext a obsah archivních zdrojů, nese administrativní informace k využívání zdroje
<archdescgrp> Archival Description Group	užívá se pouze v DTD pro popis skupiny zdrojů, popisuje kontext a obsah archivních materiálů
<archref> Archival Reference	citace či přímo elektronický link na samostatně popisovaný zdroj
<arrangement> Arrangement	struktura a uspořádání popisovaného materiálu
<author> Author	údaje o autorovi (osobě či instituci)
<bibliography> Bibliography	doplňující prvek pro záznam materiálů použitých při tvorbě zdroje, prvek může být začleněn do dalších prvků
<bibref> Bibliographic Reference	citace nebo elektronický odkaz na publikované dílo
<bibseries> Bibliographic Series	informace o edici, v jejímž rámci byla publikována popisovaná jednotka
<bioghist> Biography or History	údaje spojené s osobou či korporací, která se podílela na vzniku zdroje
<blockquote> Block Quote	formátovací prvek označující citovaný text
<c> Component	prvek sdružující prvky o obsahu a kontextu části zdroje, může obsahovat další podřízené číslované prvky
<c01> Component	podřízený prvek prvku <c>, první úroveň
<c02> Component	podřízený prvek prvku <c>, druhá úroveň
<c03> Component	podřízený prvek prvku <c>, třetí úroveň
<c04> Component	podřízený prvek prvku <c>, čtvrtá úroveň
<c05> Component	podřízený prvek prvku <c>, pátá úroveň
<c06> Component	podřízený prvek prvku <c>, šestá úroveň
<c07> Component	podřízený prvek prvku <c>, sedmá úroveň
<c08> Component	podřízený prvek prvku <c>, osmá úroveň
<c09> Component	podřízený prvek prvku <c>, devátá úroveň
<c10> Component	podřízený prvek prvku <c>, desátá úroveň
<c11> Component	podřízený prvek prvku <c>, jedenáctá úroveň
<c12> Component	podřízený prvek prvku <c>, dvanáctá úroveň
<colspec> Table Column Specification	formátovací prvek určující pozici a velikost sloupce v tabulce
<container> Container	lokační prvek obsahující informace o typu nosiče zdroje
<controlaccess> Controlled Access Headings	prvek určující přístupový bod ke zdroji
<corpname> Corporate Name	identifikace organizace či skupiny jako entity

Prvek	Popis
<creation> Creation	podřízený prvek prvku <profiledesc> nesoucí informace o kódování zdroje a autoritách odpovědných za toto kódování
<custodhist> Custodial History	prvek vyššího prvku <admininfo> s informacemi o historii vlastnických práv zdroje
<dao> Digital Archival Object	prázdný prvek, který propojuje přes vlastnosti ENTITYREF nebo XLINK záznam s popisovaným materiálem
<daodesc> Digital Archival Object Description	informace o obsahu, užití nebo zdroji prvků <dao> a <daogrp>
<daogrp> Digital Archival Object Group	prvek obsahující prvky <daoloc> a <daodesc>
<daoloc> Digital Archival Object Location	určení lokace zdroje, je součástí prvku <daogrp>
<date> Date	prvek obsahující měsíc, den a rok v jakémkoliv formátu, datum se může vázat ke zdroji i autoritě
<defitem> Definition List Item	formátovací prvek určený k označení položky seznamu
<descgrp> Description Group	sruzuje jakékoliv dceřinné prvky v prvku <archdesc>
<descrules> Descriptive Rules	slouží pro výčet použitých pravidel a standardů při tvorbě popisu
<did> Descriptive Identification	prvek vážící hlavní informace o zdroji
<dimensions> Dimensions	součást prvku <physdescs>, nese informace o velikosti zdroje
<div> Text Division	prvek obsahující údaje z titulní stránky, předmluvu, poděkování, informace pro užití apod., součást prvku <frontmatter>
<dsc> Description of Subordinate Components	prvek vážící informace o hierarchické struktuře a seskupování popisovaných materiálů
<dscgrp> Description of Subordinate Components Group	prvek užívaný pouze v rámci prvku <archdescgrp>, obsahuje kódované prvky <ead>
<ead> Encoded Archival Description	nejvyšší prvek určující, že se jedná o popis EAD
<eadgrp> EAD Group	jeden z nejvyšších prvků popisu EAD, obsahuje vnořené prvky
<eadheader> EAD Header	hlavička EAD
<eadid> EAD Identifier	identifikátor užitý v hlavičce EAD
<edition> Edition	označení verze zdroje
<editionstmt> Edition Statement	souborné informace o verzi zdroje uložené v hlavičce EAD
<emph> Emphasis	význam zdroje
<entry> Table Entry	formátovací prvek určující obsah buňky v tabulce
<event> Event	jeden prvek vyššího prvku <chronitem>, popisuje nějakou událost
<eventgrp> Event Group	prvek sdružující více událostí vážících se ke stejnému datu
<expan> Expansion	plná verze zkratky či akronymu z prvku <abbr>
<extent> Extent	informace o množství popisovaných materiálů nebo velikosti jednoho zdroje
<extptr> Extended Pointer	prvek odkazující na externí zdroj
<extptrloc> Extended Pointer Location	lokace odkazu na externí zdroj
<extref> Extended Reference	propojovací prvek, který může nést text nebo další prvky s odkazem na externí zdroj
<extrefloc> Extended Reference Location	lokace rozšířené citace zdroje
<familyname> Family Name	příjmení
<filedesc> File Description	sduzuje prvky s bibliografickým popisem jednotky
<fileplan> File Plan	informace o použitých řízených slovnících
<frontmatter> Front Matter	informace předcházející samotnému popisu jednotky, jde např. o údaje o vzniku či publikování zdroje
<function> Function	informace o procesech či událostech vedoucích ke vzniku materiálu

Prvek	Popis
<genreform> Genre/Physical Characteristic	žánr či fyzická forma, v jaké zdroj vznikl (jde o použité techniky nebo fyzické charakteristiky zdroje)
<geogname> Geographic Name	určení místa či státní správní jednotky
<head> Heading	záhlaví (název či název kapitoly či části zdroje), může obsahovat hierarchický seznam dalších záhlaví
<head01> First Heading	první záhlaví
<head02> Second Heading	druhé záhlaví
<change> Change	záznam pro jednotlivou úpravu či aktualizaci zdroje
<chronitem> Chronology List Item	formátovací prvek pro označení období uvedeného v prvku <chronlist>
<chronlist> Chronology List	záznam časového období, ve kterém se udála významná akce se vztahem k popisovanému zdroji
<imprint> Imprint	informace spojené s vydáním či distribucí citovaného zdroje
<index> Index	rejstřík či abecedně seřazený přehled souvisejících slov
<indexentry> Index Entry	formátovací prvek párující termín z rejstříku s propojovacím prvkem
<item> Item	označení položky v prvcích <item>, <defitem> a <change>
<langmaterial> Language of the Material	výčet jazyků, které jsou obsažené v textu popisovaného zdroje (volný text)
<language> Language	jazyk, ve kterém je zdroj napsán
<langusage> Language Usage	doplňkové informace o jazyku zdroje
<lb> Line Break	formátovací prvek, ukončuje řádku, následující text začíná na řádce další
<legalstatus> Legal Status	legální status popisovaného zdroje
<linkgrp> Linking Group	prvek sdružující prvky s odkazy
<list> List	formátovací prvek označující položku seznamu
<listhead> List Heading	formátovací prvek sdružující záhlaví rejstříků, seznamů slov a chronologicky řazených seznamů událostí
<materialspect> Material Specific Details	specifická forma informací, např. matematická nebo kartografická data
<name> Name	obecné jméno, které nelze specifikovat blíže pomocí prvků <corpname>, <famname>, <geogname>, <persname> nebo <title>
<namegrp> Name Group	formátovací prvek sdružující prvky <ref>, <ptr> a <ptrgrp>
<note> Note	všeobecný prvek obsahující krátké vysvětlení
<notestmt> Note Statement	prvek sdružující prvky <note>
<num> Number	všeobecný prvek s číselným určením
<occupation> Occupation	typ práce, profese, obchodu či zájmu, který se odráží v popisovaném zdroji
<odd> Other Descriptive Data	prvek sdružující prvky s popisem zdroje, které nelze zařadit pod prvky <archdesc> a <c>
<originalsloc> Location of Originals	informace o případných původních částech popisovaného zdroje
<origination> Origination	informace o osobní či korporativní odpovědnosti za vznik či shromáždění popisovaných zdrojů
<otherfindaid> Other Finding Aid	informace o jiných pomůckách pro popis a vyhledávání popisovaného zdroje (např. katalog vydavatele, seznam vytvořený autorem, apod.)
<p> Paragraph	jedna nebo více vět tvořící souvislý text - odstavec
<persname> Personal Name	osobní jméno zahrnující všechny části (křestní, příjmení, tituly apod.)
<physdesc> Physical Description	souborný prvek sdružující informace o fyzickém popisu zdroje
<physfacet> Physical Facet	informace o vzhledu zdroje (barva, styl, technika vzniku apod.)
<physloc> Physical Location	lokace uložení zdroje (budova, místnost, regál apod.)

Prvek	Popis
<phystech> Physical Characteristics and Technical Requirements	podmínky pro uchovávání, ochranu nebo použití popisovaného zdroje
<prefercite> Preferred Citation	podřazený prvek prvku <admininfo> s doporučenou citací popisovaného zdroje
<processinfo> Processing Information	informace spojené s uložením, popisem a úpravami materiálu
<profiledesc> Profile Description	prvek vážící informace o způsobu kódování zdroje, verzi a datu kódování
<ptr> Pointer	prázdný propojovací prvek odkazující na jinou část téhož zdroje, prvek neobsahuje text
<ptrgrp> Pointer Group	prvek sdružující více prvků <ptr> a <ref>
<ptrloc> Pointer Location	lokace prvku <ptr>
<publicationstmt> Publication Statement	prvek sdružující informace o vydání zdroje (nakladatelské údaje, datum vydání apod.)
<publisher> Publisher	nakladatelské údaje
<ref> Reference	propojovací prvek s odkazem na jinou část téhož zdroje, může obsahovat text
<refloc> Reference Location	určení lokace prvku <ref>
<relatedmaterial> Related Material	informace o materiálech, které jsou asociované s popisovaným zdrojem
<repository> Repository	instituce, která je odpovědná za zpřístupnění zdroje
<resource> Resource	propojovací prvek na lokální zdroj
<revisiondesc> Revision Description	historie změn popisu zdroje
<row> Table Row	formátovací prvek pro označení řádky tabulky
<runner> Runner	formátovací prvek pro záhlaví, zápatí či vodoznak
<scopecontent> Scope and Content	prvek shrnující oborové pokrytí zdroje
<separatedmaterial> Separated Material	informace o materiálu, který má stejný původ jako popisovaný zdroj, ale je od zdroje fyzicky oddělen
<seriesstmt> Series Statement	prvek sdružující informace o edici, která má vztah k popisovaným zdrojům
<sponsor> Sponsor	informace o sponzorovi (osoba či instituce)
<subarea> Subordinate Area	určení hierarchické úrovně archivu, ve které je uložen popisovaný zdroj
<subject> Subject	předmět zdroje
<subtitle> Subtitle	podnázev
<table> Table	formátovací prvek pro označení tabulky
<tbody> Table Body	formátovací prvek obsahující jednu či více řádek tabulky
<tgroup> Table Group	formátovací prvek vážící prvky tabulky <colspec>, <spanspec>, <thead>, <tbody> a <tfoot>
<thead> Table Head	formátovací prvek s údaji pro záhlaví tabulky (obvykle názvy sloupců apod.)
<title> Title	formální název díla
<titlepage> Title Page	souborný prvek shrnující bibliografické informace o zdroji (název, autor, nakladatelské údaje), součást prvku <frontmatter>
<titleproper> Title Proper of the Finding Aid	název vyhledávacího nástroje, je vyžadován v prvku <titlestmt>, který je součástí prvků <filedesc> a <eadheader>
<titlestmt> Title Statement	prvek sdružující informace o názvech, je zařazen do hierarchie prvků pod <filedesc> a <eadheader>
<unitdate> Date of the Unit	datum vytvoření zdroje
<unitid> ID of the Unit	kontrolní číslo popisovaného zdroje
<unittitle> Title of the Unit	název popisovaného materiálu
<userrestrict> Restrictions on Use	okolností omezujících užití zdroje (omezení, regulace, právní stav apod.)

(EAD, 2002d)

Příloha 10: Záznam z Centrály archivů ve formátu EAD

```
<!DOCTYPE EAD PUBLIC "-//Society of American Archivists//DTD
ead.dtd (Encoded Archival Description (EAD) Version 1.0)//EN">
<!--Prepared using the Archives Hub ISAD(G) template version 2, from the University of Liverpool
Library-->
<ead>
<eadheader>
<eadid type="JISC-HUB">GB 0738 MS 1</eadid>
<filedesc>
<titlestmt><titleproper encodinganalog="">Records of the University
of Southampton</titleproper></titlestmt>
</filedesc>
<profiledesc>
<creation encodinganalog = "JISC-HUB24">This document was generated by Javascript from
an HTML form which structured the input according to the elements of ISAD(G) Version 2.
</creation>
</profiledesc>
<revisiondesc>
<list>
<item> Modified at: <date> Tue May 8 14:17:47 UTC+0100 2001</date></item></list>
</revisiondesc>
</eadheader>
<archdesc encodinganalog = "JISC-HUB04" level="fonds" langmaterial = "eng">
<did>
<unitid encodinganalog="JISC-HUB01">GB 0738 MS 1</unitid>
<unittitle encodinganalog="JISC-HUB02">Records of the University of Southampton</unittitle>
<unitdate encodinganalog="JISC-HUB03" normal = "1862xxxx-1986xxxx">1862-1986</unitdate>
<repository>University of Southampton</repository>
<physdesc>
<extent encodinganalog = "JISC-HUB05">680 boxes</extent>
<emph altrender = "JISC-HUB16">English</emph>
</physdesc>
</did>
<bioghist encodinganalog="JISC-HUB07">
<p>In 1850, Henry Robinson Hartley, a descendant of two generations of Southampton wine
merchants, bequeathed nearly the whole of his estate, amounting to more than 100,000 li., to the
Corporation of Southampton to be employed 'in such manner as [might] best promote the study
and advancement of the sciences of natural history, astronomy, antiquities, classical and oriental
literature in the town, such as by forming a public library, botanic gardens, observatory,
and collections of objects in connection with the above sciences'. The will was contested and the
terms of the benefaction, which were ambiguous in some degree, made the task of the Corporation
more difficult; and it was not until 1862 that the Prime Minister, the third Viscount Palmerston, was
able to open the newly built Hartley Institution, with its library, museum and lecture halls that had
been constructed on the site of three houses leased by H.R.Hartley in the High Street in Southampton.
</p><p>
Between 1863 and the end of the nineteenth century, instruction was offered in a range of scientific,
technical, engineering and literary subjects, and the Southampton School of Art was subsumed within
the Hartley Institution in 1871. The museum of the Institution was open to the public on three days
of each week and a number of important exhibitions was staged. Aside from its endowment, the
Institution was partly subvented by Southampton Corporation, in particular for 'technical education'.
</p><p>
The Institution was incorporated on 23 November 1902 as the Hartley University College, the result
of recognition by the University Colleges' Grants Commission of the considerable development
and achievement in the Institution's teaching of the sciences and the humanities and the campaign
for a university for Wessex. The activities of the College continued to expand and while limited
additions could and were made to its accommodation in the town, a new site with potential for greater
expansion was required. Land in Highfield was acquired but one of the effects of the First World War
was to postpone for five years the move of the University College of Southampton, as it became
```

in 1914, to its new site. For the duration of the war the new buildings in Highfield were used as a war hospital and Highfield Hall was lent to the Red Cross Society.

</p><p>

The College continued to expand, with the construction of further buildings and a very considerable increase in the number of students, particularly after the Second World War. The University College became the University of Southampton with the grant of a royal charter on 29 April 1952, with the seventh Duke of Wellington as its first Chancellor. In 1989-90, it had more than 7,500 students and approximately 1,615 academic and academically-related staff, divided into seven faculties: arts, social sciences, law, engineering, science, mathematics and medicine.</p>

</bioghist>

<scopecontent encodinganalog="JISC-HUB10">

<p>Minute books: Council, 1869-1968; Senate, 1906-65; Court of Governors, 1903-51, with registers of members of Council and Court and of attendance, 1902-51; general purposes and financecommittee, 1872-1968; Hartley bequest committee, 1860-8. Records of other committees including: apparatus grant committee, 1949-64; development committee, 1908-40; educational committee, 1894-1906; emergency executive committee, 1939-45; grants committee, 1919-20; grounds committee, 1930-48; halls of residence committee, 1930-52; laboratory technical staff committee, 1959-67; library committee, 1915-58; union facilities committee, 1964-7; works committee, 1938-52. Minute books for the Hartley Institution Library and Museum, 1872-98; Hampshire Loan Exhibition, 1866.

</p><p>

Accounts and ledgers, 1912-43; accounts - hostels, 1908-17.

</p><p>

Registers of contracts of appointment, 1892-1945

</p><p>

Letter books, 1894-8, 1907-15

</p><p>

Records of the University's central administration: selected files of the Secretary and Registrar and of the Academic Registrar from 1952 onwards. Some records from faculties, especially Arts, Education and Engineering.

</p><p>

Papers of societies, including the Hartley University College Past Students' Association, later the Hartley Society, 1891-1964, the Engineering Faculty Society, 1956-71.

</p>

</scopecontent>

<admininfo>

<accessrestrict encodinganalog="JISC-HUB14">Open.</accessrestrict>

</admininfo>

<add>

<otherfindaid encodinganalog="JISC-HUB18"><p>A summary description of the archives of the University and its predecessors is available in G.Hampson <title>Records of the University of Southampton</title> (University of Southampton Library Occasional Paper, 7; 1980). </p></otherfindaid>

<relatedmaterial encodinganalog="JISC-HUB21">

<p>Most of the archives of the Hartley Institution, the University College of Southampton and the University of Southampton are now held together in one collection in the University Library, although some early records form a part of the archives of the Corporation of Southampton and are preserved in the Southampton City Record Office. The papers of former students and staff are maintained as separate collections.</p>

</relatedmaterial>

<bibliography encodinganalog="JISC-HUB22">

<p>The history of the institution was written by A.Temple Patterson

<title>The University of Southampton</title> (University of Southampton, Southampton, 1962); and there is a study of H.R.Hartley by A.Anderson <title>Hartleyana, being some account of the life and opinions of Henry Robinson Hartley </title> (Southampton Record Series,supplementary volume, 1987).</p>

</bibliography>

</add>

```

<!-- Control Access Terms -->
<controlaccess>
<subject encodinganalog = "JISC-HUB43"source = "LCSH">
<emph altrender = "a">Education, Higher</emph>
<emph altrender = "z">Great Britain</emph>
</subject>
<subject encodinganalog = "JISC-HUB43"source = "LCSH">
<emph altrender = "a">Humanities</emph>
</subject>
<persname encodinganalog = "JISC-HUB40">
<emph altrender= "surname">Hartley</emph>
<emph altrender= "forename">Henry Robinson</emph>
<emph altrender= "dates">1777-1850</emph>
<emph altrender= "epithet">wine merchant; scholar; founder of the Hartley Institution</emph>
</persname>
<persname encodinganalog = "JISC-HUB40">
<emph altrender= "surname">Wellesley</emph>
<emph altrender= "forename">Gerald</emph>
<emph altrender= "dates">1885-1972</emph>
<emph altrender= "title">7th Duke of Wellington</emph>
<emph altrender= "epithet">first chancellor of the University of Southampton</emph>
</persname>
<corpname encodinganalog = "JISC-HUB41">
<emph altrender = "a">Hartley Institution</emph>
<emph altrender = "z">Southampton</emph>
</corpname>
<corpname encodinganalog = "JISC-HUB41">
<emph altrender = "a">University of Southampton</emph>
</corpname>
<geogname encodinganalog = "JISC-HUB42">
<emph altrender = "a">Southampton (England)</emph>
</geogname>
</controlaccess>
</archdesc>
</ead>

```

(Archives Hub, 2000e)

Příloha 11: Zobrazení vyhledného záznamu z databáze Janus

Papers of Dawson Turner (1775-1858), banker, botanist and antiquary

Title Papers of Dawson Turner (1775-1858), banker, botanist and antiquary
Reference GBR/0016/TURNER1
Creator Dawson Turner
Covering Dates 1790-1857
Extent and Medium 83 volumes
Repository Trinity College Library, Cambridge

Content and context

Dawson Turner was born on 18 October 1775 in Great Yarmouth, the son of James Turner, banker, and his wife Elizabeth Cotman. In 1793 he came up to Cambridge and entered Pembroke College where his uncle was Master. However, he did not graduate and left the University the following year joining his father's bank in 1796. Supported by his inheritance, Turner was able to indulge his interests. His early passion was botany with special reference to cryptogams, publishing a number of books on the subject. Throughout his life he corresponded with many of the great botanists of the day. His further two great passions were painting - his wife and children took lessons from John Sell Cotman - and antiquarian studies, pursuing the latter subject with some vigour from 1820 onwards. He died in 1858.

In-letters, 1790-1857. Correspondents include Sir Joseph Banks, William Borrer, John Sell Cotman, W J Hooker, Joseph Hume, Sir Francis Palgrave and William Upcott.

These volumes were given to Trinity College in 1890 by Ellen Jacobsen, widow of the Bishop of Chester, Turner's last surviving daughter.

Turner collected and bound together his incoming letters in a number of series. These volumes form the series of "general correspondence" and are arranged chronologically.

Access and Use

For permission to publish materials from the Dawson Turner papers please contact The Librarian, Trinity College Library, Cambridge, CB2 1TQ

Please cite as Trinity College Library, Cambridge: Dawson Turner Papers.

Further information

For a short biography of Turner see Warren R Dawson, "Dawson Turner FRS" in the Journal of the Society for the Bibliography of Natural History, vol 3:6, Dec 1958.

A contemporary personal name index of this correspondence is available and is the prime finding aid.

Trinity College Library holds microfilms of these volumes for reprographic purposes

Index Terms

[Turner, Dawson \(1775-1858\) banker, botanist and antiquary](#)

Trinity/TURNER1 contains:

O.13	Correspondence. 32 vol.	1790-Jun 1827
O.14	Correspondence. 51 vol.	Jul 1827-1851

Evidence výpůjček

Prohlášení:

Dávám svolení k půjčování této rigorózní práce. Uživatel potvrzuje svým podpisem, že bude tuto práci řádně citovat v seznamu použité literatury.

V Mratíně, 7.9.2007

Veronika Synková

[illegible]